

**STIC-Biotech/CheMlib**

137546

mej

**From:** Whiteman, Brian  
**Sent:** Wednesday, November 10, 2004 10:02 AM  
**To:** STIC-Biotech/CheMlib  
**Subject:** sequence search

10/069,386, Alaoui-Jamali et al.  
2/19/02

RECEIVED  
NOV 10 2004

search SEQ ID NO: 2 against databases.

Thank you,

Brian Whiteman  
Remsen, 2D14  
mail box 2C18  
Patent Examiner - Art Unit 1635  
United States Patent and Trademark Office  
(571) 272-0764

\*\*\*\*\*

STAFF USE ONLY

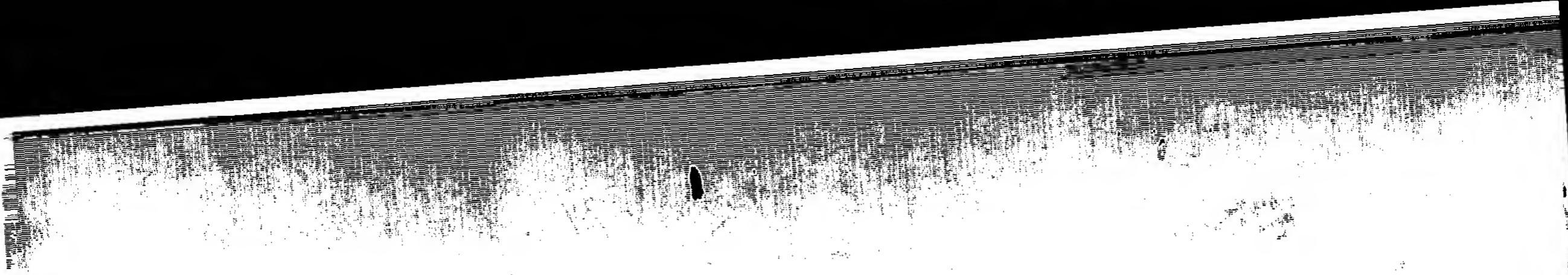
Searcher: \_\_\_\_\_  
Searcher Phone: 2-  
Date Searcher Picked up: \_\_\_\_\_  
Date Completed: \_\_\_\_\_  
Searcher Prep/Rev. Time: \_\_\_\_\_  
Online Time: \_\_\_\_\_

\*\*\*\*\*

Type of Search  
NA Sequence: # \_\_\_\_\_  
AA Sequence :# \_\_\_\_\_  
Structure: # \_\_\_\_\_  
Bibliographic: \_\_\_\_\_  
Litigation: \_\_\_\_\_  
Patent Family: \_\_\_\_\_  
Other: \_\_\_\_\_

\*\*\*\*\*

Vendors and cost where applicable  
STN: \_\_\_\_\_  
DIALOG: \_\_\_\_\_  
QUESTEL/ORBIT: \_\_\_\_\_  
LEXIS/NEXIS: \_\_\_\_\_  
SEQUENCE SYSTEM: \_\_\_\_\_  
WWW/Internet: \_\_\_\_\_  
Other(Specify): \_\_\_\_\_





QY 1 MEGGLKRKHSDELLEEEERNEWSPAGLQSYQQALLRISLDKVQRSLGPRAPSRRHVLHN 60  
 1 MEGGLKRKHSDELLEEEERNEWSPAGLQSYQQALLRISLDKVQRSLGPRAPSRRHVLHN 60  
 Db 61 TLQQLQAALRLAPAPALPPEPLGEEDFLSATIGSILRELDTSMDGTEPQNPNPVTPLG 120  
 "DB 61 TLQQLQAALRLAPAPALPPEPLGEEDFLSATIGSILRELDTSMDGTEPQNPNPVTPLG 120  
 QY 121 LQNEVPPQDPFVFLAEALSSRYLGDSGLDDFFLDIDTSAVEKEPARAPEPPHNLCAPGS 180  
 DB 121 LQNEVPPQDPFVFLAEALSSRYLGDSGLDDFFLDIDTSAVEKEPARAPEPPHNLCAPGS 180  
 QY 181 WENNLDHIMEIILGS 196  
 DB 181 WENNLDHIMEIILGS 196

## RESULT 2

AAM93922

ID AAM93922 standard; protein; 196 AA.

XX AC AAM93922;

XX DE 06-NOV-2001 (First entry)

XX DE Human polypeptide, SEQ ID NO: 4085.

XX AR Human; full length cDNA; CDNA synthesis; oligo-capping.

XX DS Homo sapiens.

XX PN EP1130094-A2.

XX DD 05-SEP-2001.

XX DD 07-JUL-2000; 2000EP-0C114089.

XX DR 08-JUL-1999; 99JP-00194486.

DR 11-JAN-2000; 2000JP-00118774.

DR 02-MAY-2000; 2000JP-00183865.

PA (HELI-) HELIX RES INST.

PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

PA Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

DR WPI; 2001-524255/58.

DR N-PSDB; AAK94884.

PS Claim 8; SEQ ID NO 4085; 1380pp + Sequence Listing; English.

PT XX

The invention relates to primers for synthesising full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have been determined. Primers for synthesising the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence is a polypeptide encoded by a full length human cDNA of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO

Sequence 196 AA;

SQ

Query Match

Best Local Similarity

Matches 195;

Conservative

0;

Mismatches

1;

Indels

0;

Gaps

0;

Score 1026;

DB 8;

Length 196;

Pred. No. 7.5e-92;

Mismatches 1;

Indels 0;

Gaps 0;

=

## RESULT 3

ADL32052

ID ADL32052 standard; protein; 196 AA.

XX AC ADL32052;

XX DT 20-MAY-2004 (first entry)

XX DE Human protein encoded by a full length cDNA clone SeqID 4085.

KW human; medicine; signal transduction; glycoprotein; transcription; KW oligo-capping method.  
 XX OS Homo sapiens.  
 XX PN EP1396543-A2.  
 XX PD 10-MAR-2004.  
 XX PF 07-JUL-2000; 2000EP-00025638.  
 XX PR 08-JUL-1999; 99JP-00194486.  
 PR 11-JAN-2000; 2000JP-00118774.  
 PR 02-MAY-2000; 2000JP-00183865.  
 PR 07-JUL-2000; 2000EP-00114089.  
 XX PA (REAS-) RES ASSOC BIOTECHNOLOGY.  
 XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 XX DR WPI; 2004-204755/20.  
 DR N-PSDB; ADL32051.

New oligonucleotide primers (830 cDNAs) useful for synthesizing full length human cDNAs.  
 XX PT SEQ ID NO 4085; 1340pp. English.  
 XX PS Example 1; SEQ ID NO 4085; 1340pp. English.

This invention relates to a novel primers useful for synthesising full length cDNA molecules that encode human proteins. Specifically, it refers to secretory or membrane proteins that are potential therapeutic agents/ target molecules in the field of medicine, and in particular genes encoding proteins that are associated with signal transduction, glycoproteins and transcription. The present invention describes a method for efficiently cloning a full length human cDNA from both the 5' and 3' ends using the oligo-capping method. This polypeptide sequence is a full length human protein of the invention.

XX Sequence 196 AA;

SQ Query Match 99.3%; Score 1026;

Best Local Similarity 99.5%; Pred. No. 7.5e-92;

Matches 195;

Conservative 0;

Mismatches 1;

Indels 0;

Gaps 0;

=

1 MEGGLKRKHSDELLEEEERNEWSPAGLQSYQQALLRISLDKVQRSLGPRAPSRRHVLHN 60

1 MEGGLKRKHSDELLEEEERNEWSPAGLQSYQQALLRISLDKVQRSLGPRAPSRRHVLHN 60

Db	1	MVGGGLRKHSDDLEEEBERNEWSAGLQSYQQALLRISLDKVQRSLGPRAPSILRRHVLHN	60	Query Match 76.6%; Score 791; DB 4; Length 157;
Qy	61	TIQQLOAALRLAPAPALPPEPLFLGEEDFSLSATIGSILRELDTSMDGTTEPPONPVTPLG	120	Best Local Similarity 98.0%; Pred. No. 5.1e-69;
Db	61	TIQQLOAALRLAPAPALPPEPLFLGEEDFSLSATIGSILRELDTSMDGTTEPPONPVTPLG	120	Matches 150; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
Qy	121	LQNBYPPQPDPVFLFLEALSSRYLGDSGLDDFFLDIDTSAVEKEPARAPPEPPHNLFCAPGS	180	Qy 44 SLGPRAPISLRRHVLIHNTLQOLOAALRLAPAPALPPEPLFLGEEDFSLSATIGSILRELD 103
Db	121	LQNEVPPQPDPVFLFLEALSSRYLGDSGLDDFFLDIDTSAVEKEPARAPPEPPHNLFCAPGS	180	Db 5 ALGPRAPISLRRHVLIHNTLQOLOAALRLAPAPALPPEPLFLGEEDFSLAXIGSILRELD 64.
Qy	181	WEWNELDHIMEIIILGS 196		Qy 104 TSMGDGTEPPQNPNVTPLGLQNEVPQPDPVFLFLEALSSRYLGDSGLDDFFLDIDTSAVEKEP 163
Db	181	WEWNELDHIMEIIILGS 196		Db 65 TSMGDGTEPPQNPNVTPLGLQNEVPQPDPVFLFLEALSSRYLGDSGLDDFFLDIDTSAVEKEP 124
Qy	164	ARAPPEPHNLFCAPGSWEWNELDHIMEIIILGS 196		Qy 164 ARAPPEPHNLFCAPGSWEWNELDHIMEIIILGS 196
Db	125	ARAPPEPXHNLFCAPGSWEWNELDHIMEIIILGS 157		Db 125 ARAPPEPXHNLFCAPGSWEWNELDHIMEIIILGS 157
<hr/>				
RESULT 4				
	AAG74742	standard; protein; 157 AA.		RESULT 5
ID	AAG74742		ID ABO60075	ABO60075 standard; protein; 142 AA.
XX	AAC		XX	ABO60075;
XX	DT	03-SEP-2001 (first entry)	XX	AC
XX	DE	Human colon cancer antigen protein SEQ ID NO:5506.	DT	29-JUL-2004 (first entry)
KW	KW	Human; colon cancer; colon cancer antigen; diagnosis; detection;	XX	DE Human genome derived single exon protein #6309.
KW	KW	colorectal carcinoma.	XX	KW Human; gene expression; single exon probe; microarray;
XX	OS	Homo sapiens.	XX	KW alternative splicing event; genomic alteration.
XX	OS		XX	XX Homo sapiens.
PN	PN	WO200122920-A2.	XX	PN US2003194704-A1.
XX	PD	05-APR-2001.	XX	PD 16-OCT-2003.
XX	PD	28-SEP-2000; 2000WO-US026524.	XX	XX 03-APR-2002; 20002US-00029386.
XX	PF	29-SEP-1999; 99US-0157137P.	XX	XX 03-APR-2002; 20002US-00029386.
PR	PR	03-NOV-1999; 99US-0163280P.	XX	XX (PENN/) PENN S G.
XX	PA	(HUMA-) HUMAN GENOME SCI INC.	XX	PA (RANK/) RANK D R.
PA	PA	Ruben SM, Barash SC, Birse CE, Rosen CA;	XX	PA (HANZ/) HANZEL D K.
XX	WPI	2001-235357/24.	XX	XX PI Penn SG, Rank DR, Hanzel DK;
DR	DR	N-PSDB; AAH34147.	XX	XX DR WPI; 2004-119264/12.
XX	PT	Nucleic acids encoding 4277 human colon cancer-associated polypeptides, useful for preventing, diagnosing and/or treating colorectal cancers.	XX	XX PT New human genome-derived single exon nucleic acid probes useful for human gene expression analysis, for identifying or characterizing alternative splicing events, for assessing genomic alterations or as tools for surveying tissues.
XX	PT	Claim 11; Page 7098-7099; 9803pp; English.	XX	XX PT XX PS Claim 45; SEQ ID NO 33709; 80pp; English.
XX	PT	AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon cancer-associated nucleic acid molecules (N) and proteins (P), where the proteins are collectively known as colon cancer antigens. The colon cancer antigens have cytotoxic activity and can be used in gene therapy and vaccine production. N and P may be used in the prevention, diagnosis and treatment of diseases associated with inappropriate P expression. For example, N and P may be used to treat disorders associated with decreased expression by rectifying mutations or deletions in a patient's genome that affect the activity of P by expressing inactive proteins or to supplement the patients own production of P. Additionally, N may be used to produce the colon cancer-associated Ps, by inserting the nucleic acids into a host cell and culturing the cell to express the proteins. N and P can be used in the prevention, diagnosis and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204 and AAB7789 represent sequences used in the exemplification of the present invention. N.B. Pages 666 to 682 and page 7053 of the sequence listing were missing at time of publication, meaning no sequences are present for SEQ ID NO:1027 to 1052, 7921 and 7922	XX	XX CC The invention relates to a nucleic acid probe for measuring human gene expression, comprising any of the 27,400 fully defined nucleotide sequences in the specification, or their complements or fragments, and encoding at least 8 amino acids of any of the 6888 amino acid sequences fully defined in the specification. The probe is a single exon probe that hybridises under high stringency conditions to a nucleic acid molecule expressed in human cells or tissues. Also included are a spatially-addressable set of single exon nucleic acid probes for measuring human gene expression (comprising a plurality of single exon nucleic acid probes cited above, where each of the plurality of probes is separately and addressably isolatable or amplifiable from the plurality), a single exon microarray for measuring human gene expression, a method of measuring human gene expression, a vector comprising the single exon probe cited above, an ORF-encoded peptide comprising at least 8 contiguous amino acids of any of the above-mentioned amino acid sequences (optionally with conservative amino acid substitutions), an isolated antibody that binds specifically to a peptide cited above, methods of selling and/or licensing single exon probes or microarrays to
XX	SQ	Sequence 157 AA;	CC	CC

a customer desiring to measure gene expression, a method of providing human gene expression data by subscription, and a computer-readable storage medium which contains a database having a plurality of records (each record including data on the expression of a single exon probe cited above). The probe, methods and apparatus are useful in gene expression analysis. The probes may be used as tools for surveying tissues to detect the presence of expressed messages that contain their specific exon, or in constructing genome-derived single exon microarrays. In addition, the probes are used in identifying and characterising alternative splicing events, in detecting and characterising gross alterations in the genomic locus that includes their exon, in assessing smaller genomic alterations, in priming the synthesis of nucleic acids, or in expressing the ORF-encoded peptide. The present sequence is a human single exon probe protein of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DocID=20030194704

JLT 6  
50074 ABO60074 standard; protein; 111 AA.

ABO60074 ;  
29-JUL-2004 (first entry)  
Human genome derived single exon protein #6308.  
Human; gene expression; single exon probe; microarray;  
alternative splicing event; genomic alteration.

Homo sapiens

16-OCT-2003.  
03-APR-2002; 2002US-00029386.  
03-APR-2002; 2002US-00029386.

(PENN/) PENN S G.  
(RANK/) RANK D R.  
(HANZL/) HANZEL D K.

New human genome-derived single exon nucleic acid probes useful for human gene expression analysis, for identifying or characterizing alternative splicing events, for assessing genomic alterations or as tools for

PS Claim 45; SEQ ID NO 33708; 80pp; English.

The invention relates to a nucleic acid probe for measuring human gene expression, comprising any of the 27,400 fully defined nucleotide sequences in the specification, or their complements or fragments, and encoding at least 8 amino acids of any of the 6888 amino acid sequences fully defined in the specification. The probe is a single exon probe that hybridises under high stringency conditions to a nucleic acid molecule expressed in human cells or tissues. Also included are a spatially-addressable set of single exon nucleic acid probes for measuring human gene expression (comprising a plurality of single exon nucleic acid probes cited above, where each of the Plurality of probes is separately and addressably isolatable or amplifiable from the plurality), a single exon microarray for measuring human gene expression, a method of measuring human gene expression, a vector comprising the single exon probe cited above, an ORF-encoded peptide comprising at least 8 contiguous amino acids of any of the above-mentioned amino acid sequences (optionally with conservative amino acid substitutions), an isolated antibody that binds specifically to a peptide cited above, methods of selling and/or licensing single exon probes or microarrays to a customer desiring to measure gene expression, a method of providing human gene expression data by subscription, and a computer-readable storage medium which contains a database having a plurality of records (each record including data on the expression of a single exon probe cited above. The probe, methods and apparatus are useful in gene expression analysis. The probes may be used as tools for surveying tissues to detect the presence of expressed messages that contain their specific exon, or in constructing genome-derived single exon microarrays. In addition, the probes are used in identifying and characterising alternative splicing events, in detecting and characterising gross alterations in the genomic locus that includes their exon, in assessing smaller genomic alterations, in priming the synthesis of nucleic acids or in expressing the ORF-encoded peptide. The present sequence is a human single exon probe protein of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from USPTO at seqdata.uspto.gov/sequence.html?DOCID=20030194704

RESULT 7  
ID AAY36004  
ID AAY36004 standard; protein; 236 AA.  
XX  
XX  
AC AAY36004;  
DT 13-SEP-1999 (first entry)  
XX  
DE Extended human secreted protein sequence, SEQ ID NO. 389.  
XX  
KW Secreted protein; human; cytokine; cellular proliferation; cell movement;  
KW cellular differentiation; immune system regulator; anti-inflammatory;  
KW haematopoiesis regulator; tissue growth regulator; tumour inhibitor;  
KW reproductive hormone regulator; chemotaxis; chemokinesis; gene therapy;  
KW genetic disease.  
XX  
OS Homo sapiens.  
XX  
PN WO9931236-A2.  
XX

PD 24-JUN-1999.  
 XX 17-DEC-1998; 98WO-IB002122.  
 XX WPI; 1999-385906/32.  
 DR N-PSDB; AAX97688.

PT New isolated human secreted proteins.  
 XX Claim 9; Page 330-331; 516pp; English.

This sequence is encoded by an extended human secreted protein coding sequence of the invention. The secreted proteins can be used in treating or controlling a variety of human conditions. The secreted proteins may act as cytokines or may affect cellular proliferation or differentiation or may act as immune system regulators, haematopoiesis regulators, tissue growth regulators, regulators of reproductive hormones or cell movement or have chemotactic/chemokinetic, receptor/ligand, anti-inflammatory or tumour inhibition activity. The DNAs can be used in forensic procedures to identify individuals or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal expression of the genes corresponding to the extended cDNAs. They are also useful for the genes corresponding to the extended cDNAs. They can constructing a high resolution map of the human chromosomes. They can also be used for gene therapy to control or treat genetic diseases.

XX Sequence 236 AA;

Query Match 20.9%; Score 216; DB 2; Length 236;  
 Best Local Similarity 32.6%; Pred. No. 1.3e-12;  
 Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;

XX SQ

```

QY 4 GLRKHKSDLEEE---ERNEWSPAGLQSYQQA-----LLRISLDKVQRSLGPRAPS 51
  5 GLRKREEEEEKEPLAVDSW-WLDPGHAAVAQAPPAVASSSLFDLSVKKLHHSLQXSXPD 63
Db 52 LRRHVLIHTNLQQALRLAPAPALPPEPL-----FILGEEDFSLSATIGSILRE 101
  64 LRHLVLVXNTLRRIQAS--MAPAAAALPPVPTPPAAPXVADNLASSDAALSASMAXILED 121
Db 102 LDTSMDGTEPPQNPPVTPGLQNEVPQPDP----PVFLEAL----SSRYLGDSGLDDFFL 152
  122 L-SHIEGLSQAPQP----LADEGGPPRSIGGXPPXLGA LDNGLEGLF 175
Db 153 DIDTSAVEKE--PARAPPEP-PHNLFCAPGSWE----WNELDHIMEIIIGS 196
  176 DIDTSMYDNEWLWAPASEGLKPGFED---GPGKEEAPELDEAELDYLMDVLYGT 225
Db 122 L-SHIEGLSQAPQP----LADEGGPPGRSIGGXPPXLGA DLLGPATGCCLIDNGLEGLF 175

```

RESULT 8  
 ADP19312 DT 26-AUG-2004 (first entry)  
 ID ADP19312 standard; protein; 236 AA.  
 DE Human secreted polypeptide #163.  
 KW Human; secreted protein; Genetic disease.  
 XX Homo sapiens.  
 OS US2004110939-A1.  
 PN

PD 10-JUN-2004.  
 XX 15-OCT-2001; 2001US-00978360.  
 PR 17-DEC-1998; 98WO-IB002122.  
 PR 09-FEB-1999; 99WO-IB000282.  
 PR 21-JUN-2000; 2000WO-IB000951.  
 PR 15-SEP-2000; 2000US-00663600.  
 XX PA (GEST ) GENSET SA.

XX PI Dumas Milne Edwards J, Bougueret L, Jobert S, Cluseel C;  
 PI Duclert A;

XX DR 2004-440404/41.  
 DR N-PSDB; ADP18907.

PT New isolated polynucleotide encoding secreted polypeptide, useful for gene therapy, or in diagnostic procedures to identify individuals having genetic diseases resulting from abnormal expression of the genes.

XX PS Claim 2; SEQ ID NO 568; 113pp; English.

The invention relates to human cDNA sequences that encode human secreted proteins. The invention also relates to an antibody that specifically binds to a polypeptide of the invention and a method of binding the polypeptide to an antibody. The polynucleotides are useful for expressing the entire secreted proteins which they encode and for distinguishing between human tissues and cells, and for distinguishing between human tissues and cells that do or do not express the polynucleotides comprising the cDNAs. The polynucleotides and polypeptides are useful in forensic procedures or diagnostic procedures to identify individuals with genetic diseases resulting from abnormal expression of the genes corresponding to the cDNAs. The sequences are also useful in gene therapy to control or treat genetic diseases. This sequence represents a human secreted polypeptide of the invention. Note: The sequence data for this patent did not form part of the printed specification but was obtained in electronic format from USPTO at seqdata.uspto.gov/sequence.html.

XX SQ Sequence 236 AA;

Query Match 20.9%; Score 216; DB 8; Length 236;  
 Best Local Similarity 32.6%; Pred. No. 1.3e-12;  
 Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;

XX SQ

```

QY 4 GLRKHKSDLEEE---ERNEWSPAGLQSYQQA-----LLRISLDKVQRSLGPRAPS 51
  5 GLRKREEEEEKEPLAVDSW-WLDPGHAAVAQAPPAVASSSLFDLSVKKLHHSLQXSXPD 63
Db 52 LRRHVLIHTNLQQALRLAPAPALPPEPL-----FILGEEDFSLSATIGSILRE 101
  64 LRHLVLVXNTLRRIQAS--MAPAAAALPPVPTPPAAPXVADNLASSDAALSASMAXILED 121
Db 102 LDTSMDGTEPPQNPPVTPGLQNEVPQPDP----PVFLEAL----SSRYLGDSGLDDFFL 152
  122 L-SHIEGLSQAPQP----LADEGGPPRSIGGXPPXLGA LDNGLEGLF 175
Db 153 DIDTSAVEKE--PARAPPEP-PHNLFCAPGSWE----WNELDHIMEIIIGS 196
  176 DIDTSMYDNEWLWAPASEGLKPGFED---GPGKEEAPELDEAELDYLMDVLYGT 225
Db 122 L-SHIEGLSQAPQP----LADEGGPPGRSIGGXPPXLGA DLLGPATGCCLIDNGLEGLF 175

```

RESULT 9  
 AAY44362 ID AAY44362 standard; protein; 236 AA.  
 AC XX AAY44362;  
 AC XX DT 14-MAR-2000 (first entry)  
 DE XX Human cell cycle regulation protein-3.

CECRP-3; cell cycle regulation protein-3; cell proliferation; cell proliferative disease; cancer; atherosclerosis; cirrhosis; hepatitis; psoriasis; immune system disorder; allergy; asthma; acquired immune deficiency syndrome; Crohn's disease; rheumatoid arthritis; gene therapy; chromosomal mapping.

Homo sapiens.

Key Location/Qualifiers  
44  
Modified-site /note= "Potential phosphorylation site"  
60  
Modified-site /note= "Potential phosphorylation site"  
73  
Modified-site /note= "Potential phosphorylation site"  
98  
Modified-site /note= "Potential phosphorylation site"  
117  
Modified-site /note= "Potential phosphorylation site"  
123  
Modified-site /note= "Potential phosphorylation site"  
180  
Modified-site /note= "Potential phosphorylation site"  
BN WO9964593-A2.  
XX  
PD 16-DEC-1999.  
XXX  
PE 08-JUN-1999; 99WO-US012906.  
XXX  
PR 08-JUN-1998; 98US-0088695P.  
XXX  
PA (INCYT-) INCYTE PHARM INC.  
PT Bandman O, Lal P, Tang YT, Corley NC, Guegler KJ, Baughn MR;  
PT Patterson C;  
DR WPI; 2000-105687/09.  
DR N-PADB; AAZ29482.

Novel regulatory proteins, for diagnosis, treatment and prevention of PT cell proliferative and immune system diseases.  
XX  
PS Claim 1; Page 70; 88pp; English.  
XX  
The present sequence is cell cycle regulation protein-3 (CECRP-3). Prints CC analytical method was used to identify this protein. CECRPs are CC activators of cell proliferation or inhibitors of cellular processes that CC modulate proliferation. They are used to treat or prevent cell CC proliferative diseases like cancers, atherosclerosis, cirrhosis, CC hepatitis, psoriasis, immune system disorders (e.g. acquired immune CC deficiency syndrome, allergy, asthma, Crohn's disease, rheumatoid CC arthritis). Antibodies are raised to screen for specific binding agents. CC The corresponding nucleic acid is used in gene therapy, chromosomal CC mapping and isolation of related sequences

Sequence 236 AA;  
Query Match 20.6%; Score 213; DB 3; Length 236;  
Best Local Similarity 32.6%; Pred. No. 2.6e-12;  
Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;  
4 GLRKHSDLEEE---ERWEWSPLAGLQSYQQA-----LLRISLDKVQRSLGPRAPS 51  
5 GLRKREEEEEKEPLAVDSW-WLDPGHAAVAQAPPVASSSLFDLSVLRKHHSLQQSEPD 63  
DB 5 GLRKHLHNTLQQQAAIRLAPAPLPPEPL-----FLGEEDFSLSATIGSILRE 101  
52 LRHVHLVNTLRRIQAS-MAPAAAALPPVPPSPPAAPSVADNLASSDAALSMASLLED 121  
64 LRHVHLVNTLRRIQAS-MAPAAAALPPVPPSPPAAPSVADNLASSDAALSMASLLED 121  
DB 52 LRRHVHLHNTLQQQAAIRLAPAPLPPEPL-----FLGEEDFSLSATIGSILRE 101  
64 LRHVHLVNTLRRIQAS-MAPAAAALPPVPPSPPAAPSVADNLASSDAALSMASLLED 121  
DB 102 LDTSMDGTEPPQNPNPVTPGLQNEVP-----QDPDVFLERAL--SSRYLGDSGLDDFFL 152  
122 L-SHIEGLSQAPQP----LADEGPGRSIGGAAPSLGALDLGGPATGCLLDGLEGLFEE 175

QY 153 DIDTSAVEKE---PARAPPEP-PHLFCAPGSWE----WNELDHMEIILGS 196  
DB 176 DIDTSMYDNEWLWAPASEGLKPGPED--GPCKEEAPELDEAELDYLMDVLTG 225

RESULT 10

AAM93724  
ID AAM93724 standard; protein; 236 AA.  
XX  
AC AAM93724;  
XX  
DT 06-NOV-2001 (first entry)  
XX  
DE Human polypeptide, SEQ ID NO: 3677.  
XX  
KW Human; Full length cDNA; cDNA synthesis; oligo-capping.  
OS Homo sapiens.  
XX  
PN EP1130094-A2.  
XX  
PD 05-SEP-2001.  
XX  
PF 07-JUL-2000; 2000EP-00114089.  
XX  
PR 08-JUL-1999; 99JP-00194486.  
PR 11-JAN-2000; 2000JP-00118774.  
PR 02-MAY-2000; 2000JP-00183765.  
XX  
PA (HELI-) HELIX RES INST.  
XX  
Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
XX  
WPI; 2001-524255/58.  
DR N-PADB; AAK94674.  
XX  
PT 830 Primers useful for synthesizing full length cDNA clones and their use  
in genetic manipulation.  
XX  
PS Claim 8; SEQ ID NO 3677; 1380pp + Sequence Listing; English.  
XX  
CC The invention relates to primers for synthesising full length cDNA  
CC clones. 830 cDNA molecules encoding a human protein have been isolated  
CC and nucleotide sequences of 5'- and 3'-ends of the cDNA molecules have  
CC been determined. Primers for synthesising the full length cDNA are useful  
CC for clarifying the function of the protein encoded by the cDNA. The full  
CC length clones were obtained by construction of full length enriched cDNA  
CC libraries that were synthesised by the oligo-capping method. The primers  
CC enable the production of the full length cDNA easily without any special  
CC methods. The present sequence is a polypeptide encoded by a full length  
CC human cDNA of the invention. Note: The sequence data for this patent did  
CC not form part of the printed specification, but was obtained in CD-ROM  
CC format directly from EPO  
XX  
SQ Sequence 236 AA;

Query Match 20.6%; Score 213; DB 4; Length 236;  
Best Local Similarity 32.6%; Pred. No. 2.6e-12;  
Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;  
4 GLRKHSDLEEE---ERWEWSPLAGLQSYQQA-----LLRISLDKVQRSLGPRAPS 51  
5 GLRKREEEEEKEPLAVDSW-WLDPGHAAVAQAPPVASSSLFDLSVLRKHHSLQQSEPD 63  
DB 5 GLRKHLHNTLQQQAAIRLAPAPLPPEPL-----FLGEEDFSLSATIGSILRE 101  
52 LRHVHLVNTLRRIQAS-MAPAAAALPPVPPSPPAAPSVADNLASSDAALSMASLLED 121  
64 LRHVHLVNTLRRIQAS-MAPAAAALPPVPPSPPAAPSVADNLASSDAALSMASLLED 121  
DB 102 LDTSMDGTEPPQNPNPVTPGLQNEVP-----QDPDVFLERAL--SSRYLGDSGLDDFFL 152  
122 L-SHIEGLSQAPQP----LADEGPGRSIGGAAPSLGALDLGGPATGCLLDGLEGLFEE 175

Qy 153 DIDTSAVEKE---PARAPEP-PHNLFCAPGSWE----WNELDHIMEIILGS 196  
 Db 176 DIDTSMYDNEWLWAPASEGLKPGPED---GPCKEEAPELDEAELDYLMDVLTG 225

RESULT 11  
 AAG89292 ID AAG89292 standard; protein; 236 AA.  
 XX AC AAG89292;  
 XX DT 11-SEP-2001 (first entry)  
 DE Human secreted protein, SEQ ID NO: 412.  
 XX KW Human; secreted protein; gene therapy; vaccine; treatment; diagnosis;  
 KW GENSET.  
 XX OS Homo sapiens.  
 XX PN WO200142451-A2.  
 XX PD 14-JUN-2001.  
 XX PF 07-DEC-2000; 2000WO-IB001938.  
 XX PR 08-DEC-1999; 99US-0169629P.  
 PR 06-MAR-2000; 2000US-0187470P.  
 XX PA (GEST ) GENSET.  
 PI Dumas Milne Edwards J, Bougueret L, Jobert S;  
 XX WPI; 2001-367870/38.  
 DR N-PSDB; AAH64895.  
 XX PT Full length GENSET human nucleic acids encoding potentially secreted  
 PT proteins, useful in gene therapy and vaccination against a variety of  
 PT diseases, and for diagnosis of those diseases.  
 XX PS Claim 21; Page 889-890; 921pp; English.  
 XX CC The invention relates to full length GENSET human nucleic acids encoding  
 potentially secreted proteins. The nucleic acids and the polypeptides  
 they encode may be used in the prevention, treatment and diagnosis of  
 diseases associated with inappropriate GENSET gene expression. For  
 example, they be used to treat disorders associated with decreased GENSET  
 gene expression by rectifying mutations or deletions in a patient's  
 genome that affect the activity of GENSET or by supplementing the  
 patients own production of GENSET polypeptides. Conversely, antisense  
 nucleic acid molecules may be administered to down regulate GENSET  
 expression by binding with the cells' own genes and preventing their  
 expression. The sense and antisense nucleic acids may also be used as DNA  
 probes in diagnostic assays to detect and quantitate the presence of  
 similar nucleic acid sequences in samples, and hence to determine which  
 patients may be in need of restorative therapy. The GENSET polypeptides  
 may be used as antigens in the production of antibodies and in assays to  
 identify modulators (agonists and antagonists) of GENSET polypeptide  
 expression and activity. The present sequence is a GENSET polypeptide of  
 the invention

XX SQ Sequence 236 AA;

Query Match 20.6%; Score 213; DB 4; Length 236;  
 Best Local Similarity 32.6%; Pred. No. 2.6e-12;  
 Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;

4 GLKRKHSDLEEE---ERNEWSPAGLQSYQQA-----LLRISLDKVQRSLGPRAPS 51  
 5 GLKRKREEEEEEKEPLAVDSW-WLDPGHTAVAQAPPVAVASSSSLFDLSVLUHLHSLOQSEPD 63

Db 52 LRRHVLINTLQQLQAAIRLAPAPALPPEPL-----FLGEDDFSLSATIGSILRE 101  
 Qy 52 LRRHVLINTLQQLQAAIRLAPAPALPPEPL-----FLGEDDFSLSATIGSILRE 101

Db 64 LRHLVUVNTLRIQAS--MAPAAALPPVSPPAAPSVADNLASSDAALSAASMASILED 121  
 Qy 102 LDTSMDGTTEPPQNPTVPLGLQNEVPP----QDPDPVFLAE--SRYLGDSLDDDFL 152  
 Db 122 1-SHIEGLSQAPQP---LADEGGPPGRSIGAAPSIGALDLIGPAGCULDDGLEGLFE 175  
 Qy 153 DIDTSAVEKE---PARAPEP-PHNLFCAPGSWE----WNELDHIMEIILGS 196  
 Qy 176 DIDTSMYDNEWLWAPASEGLKPGPED---GPCKEEAPELDEAELDYLMDVLTG 225

RESULT 12  
 ADL31644 ID ADL31644 Standard; protein; 236 AA.  
 XX AC ADL31644;  
 XX DT 20-MAY-2004 (first entry)  
 DE Human protein encoded by a full length cDNA clone SeqID 3677.  
 XX KW human; medicine; signal transduction; glycoprotein; transcription;  
 KW oligo-capping method.  
 XX OS Homo sapiens.  
 XX PN EP1396543-A2.  
 XX PD 10-MAR-2004.  
 XX PF 07-JUL-2000; 2003EP-00025638.  
 XX PR 08-JUL-1999; 99JP-00194486.  
 PR 11-JAN-2000; 20000JP-00118774.  
 PR 02-MAY-2000; 20000JP-00183865.  
 PR 07-JUL-2000; 20000EP-00114089.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;  
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;  
 PI WPI; 2004-204755/20.  
 DR N-PSDB; ADL31643.

XX New Oligonucleotide primers (830 cDNAs) useful for synthesizing full  
 PT length human cDNAs.

XX Example 1; SEQ ID NO 3677; 1340pp; English.

XX This invention relates to a novel primers useful for synthesising full  
 CC length cDNA molecules that encode human proteins. Specifically, it refers  
 CC to secretory or membrane proteins that are potential therapeutic agents/  
 CC target molecules in the field of medicine, and in particular genes  
 CC encoding proteins that are associated with signal transduction,  
 CC glycoproteins and transcription. The present invention describes a method  
 CC for efficiently cloning a full length human cDNA from both the 5' and 3'  
 CC ends using the oligo-capping method. This polypeptide sequence is a full  
 CC length human protein of the invention.

XX SQ Sequence 236 AA;

Query Match 20.6%; Score 213; DB 8; Length 236;  
 Best Local Similarity 32.6%; Pred. No. 2.6e-12;  
 Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;

Qy 4 GLKRKHSDLEEE---ERNEWSPAGLQSYQQA-----LLRISLDKVQRSLGPRAPS 51  
 Db 5 GLKRKREEEEEEKEPLAVDSW-WLDPGHTAVAQAPPVAVASSSSLFDLSVLUHLHSLOQSEPD 63

64 LRHLVLVNTLRRQAS--MAPAAAALPPVSPPPAAPSVAAPSVADNLASSDAALSAMASLLIED 121

102 LDTSMDGTEPPQNPTPLGLQNEVPP-----QPDPFVFLAEI--SSRYLGDSLDDFFL 152  
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |  
 122 L-SHIEGLSQAPQP----LADEGPPGRSIGGAAPSLGALDLGPATGCLLDGLEGLF 175

153 DIDTSAVERE---PARAPEP-PHNLFCAPGSWE----WNELDHIMEIIILGS 196  
   | | | : | | : | | : | | : | | : | | : | : | : | : | : | : | : | : | : |  
 176 DIDTSMYDNEWLWAPASEGLKPGPED---GP GKKEEAPELDEAELDYLMDDVLVGT 225

SULT 13  
 AAM25550 standard; protein; 278 AA.

AAM25550;

16-OCT-2001 (first entry)

Human protein sequence SEQ ID NO:1065.

Human; cancer; ulcer; HIV infection; human immunodeficiency virus; antiinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; anaemia; antiaggregant; haemostatic; vulnerary; antiulcer; antianaemic; osteopathic; dermatological; antiallergic; antiasthmatic; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antiparkinsonian; infection; immunostimulant; gene therapy; antisense therapy; vaccine; inflammation; antianaphylactic; rheumatoid arthritis; septic shock; pancreatic dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity; genetic disease; haematopoietic disorder; platelet disorder; asthma; thrombocytopaenia; osteoporosis; severe combined immunodeficiency; allergic rhinitis; diabetes; multiple sclerosis; depression; Alzheimer's disease; Parkinson's disease; neurodegenerative disorder; neurological disorder.

Homo sapiens.

WO200153455-A2.

26-JUL-2001.

22-DEC-2000; 2000WO-US035017.

23-DEC-1999; 99US-00471275.  
 21-JAN-2000; 2000US-00488725.  
 25-APR-2000; 2000US-00552317.

(HYSEQ-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT;  
 WPI; 2001-457603/49.  
 N-PSDB; AAH99491.

Isolated human polynucleotides encoding polypeptides, useful for the treatment and diagnosis of e.g. cancer, ulcers and HIV infection.

Claim 20; Page 214; 1217PP; English.

AAH99166 to AAH9904 encode the human proteins given in AAM25225 to AAM25963. The proteins can have activities based on the tissues and cells they are expressed in, such as: antiinflammatory; antirheumatic; antiarthritic; immunosuppressive; antibacterial; endocrine; cardiant; central nervous system; virucide; anti-HIV; fungicide; antimutagen; cardiovascular; antianaemic; vulnerary; antiulcer; haemostatic; antiallergic; dermatological; antidiabetic; cytostatic; neuroprotective; antidepressant; nootropic; antidePRESSANT; antiasthmatic; antiparkinsonian; and immunostimulant. The proteins and polynucleotides encoding them can be used in gene therapy, antisense therapy and vaccine production. The proteins and polynucleotides are useful for screening for agonists or antagonists of a protein and for the treatment and diagnosis

XX XX 18-MAR-1999; 99WO-US005939.  
 XX PF 20-MAR-1998; 98US-0078803P.  
 PR 17-MAR-1999; 99US-00078803.  
 XX PA (GEMY ) GENETICS INST INC.  
 XX PI Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ;  
 XX DR WPI; 1999-562059/47.  
 XX DR N-PSDB; AAZ19894.  
 XX PT New polynucleotides derived from murine fetal cell cDNA libraries,  
 PT potentially used as, e.g. vaccines.  
 XX PS Claim 13 (a) ; Page 94; 107pp; English.  
 XX This is the predicted amino acid sequence of a novel human secreted  
 CC protein, g211, as deduced from an isolated adult blood cDNA clone (see  
 CC AAZ19894). The invention provides new human secreted proteins (see  
 CC AAY31828-38) and polynucleotides (see AAZ19893-901) isolated from foetal  
 CC cell, adult blood, adult brain and foetal kidney cDNA libraries. They are  
 CC predicted to have biological activities which would make them suitable  
 CC for treating, preventing or ameliorating medical conditions in humans and  
 CC animals, although no supporting data are given. Suggested activities  
 CC include nutritional, cytokine, tissue growth, cell proliferation and  
 CC differentiation, immunostimulant (e.g. as vaccine), immunosuppressive,  
 CC haematopoiesis regulating, activin or inhibin, chemotactic or  
 CC chemokinetic, haemostatic or thrombolytic, receptor/ligand activity,  
 CC antiinflammatory, cadherin or tumour invasion suppressor, and tumour  
 CC inhibition activities  
 XX SQ Sequence 236 AA;  
 Query Match 20.4%; Score 211; DB 2; Length 236;  
 Best Local Similarity 32.6%; Pred. No. 4.1e-12;  
 Matches 76; Conservative 35; Mismatches 70; Indels 52; Gaps 13;  
 Qy 4 GLKRKHSDDLEEE---ERNEWSPAGLQSYQQA----LIRISLDKVORSLGPRAPS 51  
 Db 5 GLKRKREEEEEKEPLAVDSW-WLDPGHAAVAQAPPVASSSLFDLSVLKLHHSLLQQSEPD 63  
 Qy 52 LRRHVLIHNTLQQLQAAIRLAPAPALPPEPL----FIEGEDFSLSATIGSILRE 101  
 Db 64 LRHLMLVNTLRIQAS--MAPAAAALPPVSPPPAAPSVAVDNLASSMASLLED 121  
 Qy 102 LDTSMDGTTEPPQNPPVTPLGLQNEVPP----QPDPVFLEAL--SSRYLGDSGLDDFFL 152  
 Db 122 L-SHIEGLSQAPQP----LAXEGPPGRSIGGAAPSLGALDLDGLEGFL 175  
 Qy 153 DIDTSAVEKE---PARAPPEP-PHNLFCAPGSWE----WNELDHIMEILLGS 196  
 Db 176 DIDTSMYDNELWAPASEGLKPGPED--GPGEKEAPELDEABLDYMDVLT 225  
 XX SQ Sequence 222 AA;  
 RESULT 15  
 ID ABO60367 standard; protein; 222 AA.  
 XX AC ABO60367;  
 XX DT 29-JUL-2004 (first entry)  
 DE Human genome derived single exon protein #6601.  
 XX KW Human; gene expression; single exon probe; microarray;  
 KW alternative splicing event; genomic alteration.  
 OS Homo sapiens.  
 XX PN US2003194704-A1.  
 PN 102 LDTSMDGTTEPPQNPPVTPLGLQNEVPP----QPDPVFLEAL--SSRYLGDSGLDDFFL 152  
 XX XX PD 16-OCT-2003.  
 XX PR 03-APR-2002; 2002US-00029386.  
 XX PR 03-APR-2002; 2002US-00029386.  
 XX PA (PENN/) PENN S G.  
 PA (RANK/) RANK D R.  
 PA (HANZ/) HANZEL D K.  
 XX PI Penn SG, Rank DR, Hanzel DK;  
 XX DR WPI; 2004-119264/12.  
 XX PS Claim 45; SEQ ID NO 34001; 80pp; English.  
 XX CC The invention relates to a nucleic acid probe for measuring human gene  
 CC expression, comprising any of the 27,400 fully defined nucleotide  
 CC sequences in the specification, or their complements, or fragments, and  
 CC encoding at least 8 amino acids of any of the 6888 amino acid sequences  
 CC fully defined in the specification. The probe is a single exon probe that  
 CC hybridises under high stringency conditions to a nucleic acid molecule  
 CC expressed in human cells or tissues. Also included are a spatially-  
 CC addressable set of single exon nucleic acid probes for measuring human  
 CC gene expression (comprising a plurality of single exon nucleic acid  
 CC probes cited above, where each of the plurality of probes is separately  
 CC and addressably isolatable or amplifiable from the plurality), a single  
 CC exon microarray for measuring human gene expression, a method of  
 CC measuring human gene expression, a vector comprising the single exon  
 CC probe cited above, an ORF-encoded peptide comprising at least 8  
 CC contiguous amino acids of any of the above-mentioned amino acid  
 CC sequences (optionally with conservative amino acid substitutions), an  
 CC isolated antibody that binds specifically to a peptide cited above,  
 CC methods of selling and/or licensing single exon probes or microarrays to  
 CC a customer desiring to measure gene expression, a method of providing  
 CC human gene expression data by subscription, and a computer-readable  
 CC storage medium which contains a database having a plurality of records  
 CC (each record including data on the expression of a single exon probe  
 CC cited above. The probe, methods and apparatus are useful in gene  
 CC expression analysis. The probes may be used as tools for surveying  
 CC tissues to detect the presence of expressed messages that contain their  
 CC specific exon, or in constructing genome-derived single exon microarrays.  
 CC In addition, the probes are used in identifying and characterising gross  
 CC alternative splicing events, in detecting and characterising gross  
 CC alterations in the genomic locus that includes their exon, in assessing  
 CC smaller genomic alterations, in priming the synthesis of nucleic acids,  
 CC or in expressing the ORF-encoded peptide. The present sequence is a human  
 CC single exon probe protein of the invention. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from USPTO at  
 CC seqdata.uspto.gov/sequence.html?DOCID=20030194704  
 XX SQ Sequence 222 AA;  
 Query Match 19.8%; Score 205; DB 8; Length 222;  
 Best Local Similarity 32.6%; Pred. No. 1.5e-11;  
 Matches 75; Conservative 34; Mismatches 69; Indels 52; Gaps 13;  
 Qy 4 GLKRKHSDDLEEE---ERNEWSPAGLQSYQQA----LLRISLDKVORSLGPRAPS 51  
 Qy 5 GLKRKREEEEEKEPLAVDSW-WLDPGHAAVAQAPPVASSSLFDLSVLKLHHSLLQQSEPD 63  
 Db 52 LRHLMLVNTLRIQAS--MAPAAAALPPVSPPPAAPSVAVDNLASSMASLLED 121  
 Qy 102 LDTSMDGTTEPPQNPPVTPLGLQNEVPP----QPDPVFLEAL--SSRYLGDSGLDDFFL 152  
 Db 122 L-SHIEGLSQAPQP----LAXEGPPGRSIGGAAPSLGALDLDGLEGFL 175  
 Qy 153 DIDTSAVEKE---PARAPPEP-PHNLFCAPGSWE----WNELDHIMEILLGS 196  
 Db 176 DIDTSMYDNELWAPASEGLKPGPED--GPGEKEAPELDEABLDYMDVLT 225  
 XX SQ Sequence 222 AA;

卷之五  
15 17:33:06 2004

122	L-SHIEGLSQAPQP-----LADEGPPGRSITGGAAAPSGLDALLGPATGCLLDGGLEGLE	175
153	DIDTSAVEKE---PARAPP-----PHNLFCAPGSWE-----WNELDHIMEII	193
176	DIDTSMYDNELWAPASEGLKPGPED---GEGKEEAPELDEAEILDYLMDVLI	222

Search completed: November 15, 2004, 14:03:14  
Job time : 159 secs

Q9EEF8 anticarsia  
 Q8naf0 homo sapien  
 Q8uzb4 grapevine f  
 Q74402 schizosacch  
 Q9c6s1 arabidopsis  
 P37370 saccharomyces  
 Q07229 saccharomyces  
 Q8bn29 mus musculus  
 Q64092 mus musculus  
 Q7tnc1 mus musculus  
 Aah63047 mus musculus  
 Q8cgd1 mus musculus  
 Q8cfq8 mus musculus  
 Q8cfpt2 mus musculus  
 98 9.5 517 2 Q9EEF8  
 98 9.5 562 2 Q8naf0  
 98 9.5 309 2 Q8uzb4  
 97.5 9.4 803 2 Q74402  
 97 9.4 1201 2 Q9c6s1  
 VRP1 YEAST  
 96 9.3 817 1 VRP1  
 96 9.3 817 2 Q07229  
 95.5 9.2 432 2 Q8bn29  
 95.5 9.2 446 1 TFE3 MOUSE  
 95.5 9.2 572 2 Q7tnc1  
 AAH63047 9.2 572 2 AAH63047  
 Q8cgd1 9.2 911 2 Q8cgd1  
 Q8cfq8 9.2 917 2 Q8cfq8  
 Q8cfpt2 9.2 1092 2 Q8cfpt2  
 32 GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.  
 DDM protein - protein search, using sw model  
 run on: November 15, 2004, 13:49:50 ; Search time 198 Seconds  
 (without alignments)  
 569.563 Million cell updates/sec  
 title: US-10-069-386-2  
 Perfect score: 1033  
 Sequence: 1 MEGGLKRKHSDELLEEEERWE.....APGSWENNEDPHMTEILIGS 196

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5  
ALIGNMENTS

Searched: 1825181 seqs, 575374646 residues  
RESULT 1

Total number of nts satisfying chosen parameters: 102318  
 Minimum DB seq length: 0  
 Maximum DB seq length: 200000000  
 ID Q9UJW9 PRELIMINARY;  
 AC Q9UJW9;  
 DT 01-MAY-2000 (TREMBUREL, 13, Created)  
 DP 01 MAY 2000 (TREMBUREL, 13, Created)

DT 01-MAR-2004 (TREMBLREL. 26, Last annotation update)  
DE RPA-binding trans-activator.  
NN

```

Homo sapiens (Human) . OS
Chordata; Craniata; Vertebrata; Euteleostomi; OC
Eukaryota; Metazoa; OC
Primates; OC
Mammalia; Eutheria; Hominidae; Homo. OX
NCBI TaxID=9606;

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is obtained by dividing the total score distribution by the total number of results.

RT "RB11, a novel transcriptional co-activator, binds the second subunit  
of Replication Protein A.";  
RT Varicella-Zoster Virus 2485/2000  
SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	DR	EMBL; AF192529;	AAF05761.1;	-.		
2	DR	InterPro; IPR009263;	SERTA.			

Best Local Similarity 100.0% ; Pred. No. 1.1e-72 ;  
 Matches 196 ; Conservative 0 ; Mismatches 0 ; Indels 0 ; Gaps 0

Qy 1 MEGGLKRKHSDELLEERWEWSAGLQQALLRISLDKVQRSLGPRAPSLLRRVLIHN 60  
Aah61808 rattus no  
Q9dcz2 mus musculu

61	TLQQQALR LAPAPAL PPEPLF GLGEED FESL SATIGS ILRELD TSM DGTE PPQN PVTPLG	120
	Q6gmb1	xenopus lae
	Q9Cwm2	mus musculu
	Qy	

61 TLQQLQAAALR LAPAPALPPEPLF LGEEEDF SLSATIGSILRELDTSMDGTEPPQNPVTPPLG 120  
Q6nxd9 brachydanio  
Aah67135 brachydan

15	125	12.1	303	2	Q'222/
16	125	12.1	314	1	STD2 HUMAN
17	115	11.1	1194	2	Q76M68

21 105.3 10.2 620 2 QSBQ18 homo sapien  
 22 105.5 10.2 760 2 Q76N32 homo sapien  
 23 105.5 10.2 760 2 BAA255508 homo sapien

				RESULTS
24	105	10.2	241	Q96CQ2
25	104.5	10.1	740	Q9UPP2
26	101.5	9.8	728	P85B HUMAN
				PRELIMINARY;
				PRT;
				196 AA.

Aah70082	101.5	9.8	728	2	AAH70082	27
Q6nvc9	100	9.7	846	2	Q6NVC9	28
Aah68186	100	9.7	846	2	AAH68186	29
Q6nxl2	98.5	9.5	314	2	Q6NXL2	30
Q96CQ2;						
DT	01-DEC-2001	(TREMBLrel.	19,	Created)		
DT	01-DEC-2001	(TREMBLrel.	19,	Last sequence update)		
DT	01-OCT-2004	(TREMBLrel.	28,	Last annotation update)		



Q9JL10; Q9DBB88;	
10-OCT-2003 (Rel. 42, Created)	
10-OCT-2003 (Rel. 42, Last sequence update)	
05-JUL-2004 (Rel. 44, Last annotation update)	
SERTA domain-containing protein 1 (Transcriptional regulator interacting with the PHD-bromodomain 1) (TRIP-Bri) (CDK4-binding protein p34SEI1) (SEI-1).	
Name=Sertadi; Synonyms=SeI1;	
Mus musculus (Mouse).	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
NCBI_TaxID=10090;	
[1] SEQUENCE FROM N.A.	
Ohtani N., Hara E.; "Cloning of mouse SEI-1 cDNA.";	
Submitted (DEC-1999) to the EMBL/GenBank/DDBJ databases.	
[2] SEQUENCE FROM N.A., FUNCTION, AND INTERACTIONS.	
MEDLINE=21231173; PubMed=11331592;	
Hsu S.-I., Yang C.M., Sim K.G., Hentschel D.M., O'Leary E., Bonventre J.V.; "TRIP-Br: a novel family of PHD zinc finger- and bromodomain-interacting proteins that regulate the transcriptional activity of E2F-1/DP-1.";	
EMBO J. 20:2273-2285 (2001).	
[3] SEQUENCE FROM N.A.	
STRAIN=C57BL/6J; TISSUE=Embryo, and Small intestine;	
MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;	
Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S., Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka T., Kiyosawa H., Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T., Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W., Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S., Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J., Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H., Nagashima T., Okido T., Pavan W.J., Pertea G., Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ringwald M., Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K., Sultan R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C., Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N., Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S., Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J., Birney E., Hayashizaki Y.; "Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs.";	
Nature 420:563-573 (2002).	
[4] SEQUENCE FROM N.A.	
MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;	
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefter C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Ustdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loqueland N.A., Peters G.J., Abramson R.D., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., McEwan P.J., Garcia A.M., Hale S., Worley K.C., Gay L.J., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,	
RX	RESULT 5
RA	STDL_HUMAN
RA	ID_STDL_HUMAN STANDARD; PRT; 236 AA.
RA	Q9UHV2; Q9BUE7;
RA	AC_10-OCT-2003 (Rel. 42, Created)
RA	DT_10-OCT-2003 (Rel. 42, Last sequence update)





[1] RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Kidney;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoaka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itch M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Ohara E., Watanuki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kiria A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer.";  
 RE Genome Res. 10:1757-1771 (2000).  
 RN [2] RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=22354683; PubMed=124666851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kiyosawa H.,  
 RA Hashimoto K.; "Isolation of full-length cDNA clones from mouse brain cDNA library  
 made by oligo-capping method"; Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.

[3] RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=22354683; PubMed=124666851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kiyosawa H.,  
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamamoto I., Gojobori T.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batyalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kainai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavani W.J., Pertea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed D.J., Reid J.C., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Kondo H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibaoka K., Shinagawa A.,  
 RA Yasumishi A., Yoshino M.,  
 RA Birney E., Hayashizaki Y.; "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs"; Nature 420:563-573 (2002).  
 RN [3]

[4] RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schulier G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Muilahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Blakesley R.W., Grimwood J.W., Green E.D., Bouffard G.G.,  
 RA Rodriguez A.C., Schmutz J., Shevchenko Y., Bouffard G.G.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalius D.E.,  
 RA Whiting M., Madan A.C., Shevchenko Y., Bouffard G.G.,  
 RA Schnarch A., Schein J.E., Jones S.J.M., Marra M.A., Dickson M.C.,  
 RA Hsu S.-I., Yang C.M., Sim K.G., Sim K.G., Hentschel D.M., O'Leary E.,  
 RA Bonvouloir J.V.; "TRIP-Br: a novel family of PHD-zinc finger- and bromodomain-  
 interacting proteins that regulate the transcriptional activity of  
 E2F-1/DP-1";  
 RN [4] RP SEQUENCE OF 129-309 FROM N.A.  
 RX MEDLINE=21231173; PubMed=11331592;  
 RA Hsu S.-I., Yang C.M., Sim K.G., Sim K.G., Hentschel D.M., O'Leary E.,  
 RA Bonvouloir J.V.; "TRIP-Br: a novel family of PHD-zinc finger- and bromodomain-  
 interacting proteins that regulate the transcriptional activity of  
 E2F-1/DP-1";  
 RN [5] RP SEQUENCE OF 129-309 FROM N.A.  
 RX MEDLINE=21231173; PubMed=11331592;  
 RA Hsu S.-I., Yang C.M., Sim K.G., Sim K.G., Hentschel D.M., O'Leary E.,  
 RA Bonvouloir J.V.; "TRIP-Br: a novel family of PHD-zinc finger- and bromodomain-  
 interacting proteins that regulate the transcriptional activity of  
 E2F-1/DP-1";  
 RN [6] RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Testis;  
 RX MEDLINE=22354683; PubMed=124666851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kiyosawa H.,  
 RA Hashimoto K.; "Isolation of full-length cDNA clones from mouse brain cDNA library  
 made by oligo-capping method"; Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.

RL EMBO J. 20:2273-2285 (2001).  
 CC -!- FUNCTION: Acts at E2F-responsive promoters to integrate signals  
 CC provided by PHD- and/or bromodomain-containing transcription  
 CC factors (By similarity).  
 CC -!- SIMILARITY: Belongs to the TRIP-Br family.  
 CC -!- SIMILARITY: Contains 1 SERTA domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC DR EMBL; AB041541; BAA95026.1; - .  
 DR EMBL; AK076787; BAC36480.1; - .  
 DR EMBL; BC014726; AAH14726.1; - .  
 DR EMBL; AF366403; AAK52832.1; - .  
 DR PIR; PT0566; PT0566.  
 DR MGD; MGI:1931026; Sertad2.  
 DR InterPro; IPR009263; SERTA.  
 DR Pfam; PF06031; SERTA; 1.  
 KW Transcription regulation.  
 FT CONFLICT 138 CTL -> FTP (in Ref. 4).  
 FT CONFLICT 142 A -> S (in Ref. 3).  
 FT CONFLICT 145 P -> S (in Ref. 4).  
 FT CONFLICT 234 T -> A (in Ref. 1).  
 SQ SEQUENCE 309 AA; 33312 MW; D4178688F0DF8F00 CRC64;  
 Query Match 13.6%; Score 140.5; DB 1; Length 309;  
 Best Local Similarity 24.8%; Pred. No. 0.0063;  
 Matches 77; Conservative 29; Mismatches 85; Indels 119; Gaps 13;  
 RN [3]

Query Match 13.2%; Score 136.5; DB 2; Length 248;  
 Best Local Similarity 30.9%; Pred. No. 0.0098;  
 Matches 58; Conservative 25; Mismatches 58; Indels 47; Gaps 8;

QY 2 EGGGRKRKHSDLEEEEEEERNNEWSPAGLQS-----YQQALLRISLDKVQRSLGPRAPSIRRH 55  
 .4 KGG-KRKFDHEDEGGLEKIVSPDGPSRVSYTLOQTIFNISLMKLYNHRPLTEPSIQLKT 62

QY 56 VLIHNTLQOQLQAAIRL-----APAPALPPPEFLIGE---- 86  
 DB 63 VLINNMMLRRIQEELKQEGSLRPAFTPSSQPSNSLSDSYQEAPPAA--PHPCDLGSTTPLE 120

QY 87 -----EDFSLSATIGSILRELDTSMDG 108  
 DB 121 ACLTPASLLEDDNDTFCTLQAHVPAAPTRLSSAALPAEKDSFSSALDEI-BELCPTSTS 179

QY 109 TEPPQN--PVTPPLGLQNEVPVQ-----PDPVFLEAL-----SRYRLGDSGLD 148  
 DB 180 TEAAHTAAPEGPKGTSSSESVQKPEGPEEGRTDDSRFMDSLPGENFEITTSGFITDLTLD 239

QY 149 D-EFLIDDTSAVEKEPA----RAPPE----PBNHLFCAPG---SWEWNEL 186  
 DB 240 DILEFADIDTSMYDFDPCTSASGTTASKMVPVSADDLKLAPYSNQVAPSQPEFKMDLTIEL 299

QY 187 DHIMEIILGS 196  
 DB 300 DHIMEVLVGS 309

RESULT 10  
 Q6GM81 ID Q6GM81 PRELIMINARY; PRT; 248 AA.  
 AC Q6GM81; PRT; 237 AA.  
 DT 05-JUL-2004 (TREMBLrel. 27, Created)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)

DE Hypothetical protein.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 OC Xenopodinae; Xenopus.  
 NCBI\_TaxID=8355; [1]

RN

SEQUENCE FROM N.A.  
 TISSUE=Kidney;  
 MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Schuler G.D., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Bhat N.K., Altenschul S.F., Zeeberg B., Buetow K.H., Schaeffer C.F., Hsieh F., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters K.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan J.A., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smalius D.E., Schnurch A., Schein J.E., Jones S.J., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences"; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RL [2]  
 RN [2]  
 SEQUENCE FROM N.A.  
 RP TISSUE=Kidney;  
 RC MEDLINE=22341132; PubMed=12454917;  
 RX Klein S.L., Strausberg R.L., Wagner L., Clifton S.W., Richardson P.; "Genetic and genomic tools for Xenopus research: The NIH Xenopus initiative"; Dev. Dyn. 225:384-391 (2002).  
 RT Dev. Dyn. 225:384-391 (2002).  
 RL RN [3]  
 SEQUENCE FROM N.A.  
 RP TISSUE=Kidney;  
 RC Klein S., Strausberg R.; Klein S.L., Strausberg (JUN-2004) to the EMBL/GenBank/DDBJ databases.  
 RA Submitted (JUN-2004); EMBL: BC074195; AAH74195.1; -.  
 DR InterPro; IPR009263; SERTA.  
 DR Pfam; PF06031; SERTA; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 248 AA; 27319 MW; E703D7E9105DF08C CRC64;  
 Query Match 17 ERNEWSPAGLQSYQQALLRISLDKVQRSLGPRAPSIRRH-----AP 73  
 Best Local Similarity 30.9%; Pred. No. 0.0098;  
 Matches 58; Conservative 25; Mismatches 58; Indels 47; Gaps 8;

QY 10 ERNEWSPAGLQSYQQALLRISLDKVQRSLGPRAPSIRRH-----AP 73  
 DB 10 ERCCSIPAIQSH--CLMNISLVLRHSLRVEPDLRFVANTLRRLQGNLQVEQCAP 67

QY 74 -----APALPPEPLFLGEEDFSLSATIGSILRELDTSMDGTEPPQ 113  
 DB 68 DMWKTSEECTRSLAVYPESKKPALENTEDPLSSMDASLYSSISTILEDLN-NFEGLS--S 124

QY 114 NPVTPLGLQNEVPQPDPVFLA-----LSRY----LGDSGLDDFFLDID 155  
 DB 125 SPLQIEDDQLCAPKANPVSGSAEDMVVKLASSSSSSSPYLLGENLQDN-LEDIFEDID 183

QY 156 TSAVEKEP 163  
 DB 184 TSMYDSDP 191

RESULT 11  
 CCA4\_MOUSE STANDARD;  
 ID CCA4\_MOUSE ID CCA4\_MOUSE STANDARD;  
 AC Q9CWM2; Q921E8; Q99MP6;  
 DT 10-OCT-2003 (Rel. 42, Created)  
 DT 10-OCT-2003 (Rel. 42, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Cell division cycle associated protein 4 (Hematopoietic progenitor protein).

GN Name=Cdca4; Synonyms=HEPP;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [11]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Bone marrow;  
 RX MEDLINE=21375891; PubMed=11482882; DOI=10.1006/bcmd.2001.0434;  
 RA Abdulla J.M., Jing X., Spassov D.S., Nachtman R.G., Jurecic R.;  
 RT "Cloning and characterization of Hepp, a novel gene expressed  
 preferentially in hematopoietic progenitors and mature blood cells.";  
 RL Blood Cells Mol. Dis. 27:667-676(2001).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Mesonephros;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamamoto I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Garibaldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Peretea G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reid J.J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,  
 RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang L.,  
 RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazumi N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 full-length cDNAs.";  
 RL Nature 420:563-573 (2002).  
 RN [13]  
 RP SEQUENCE OF 13-237 FROM N.A.  
 RC TISSUE=Breast tumor;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuller G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahay J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 CC "-!- FUNCTION: May be involved in molecular regulation of hematopoietic  
 stem cells and progenitor cell lineage commitment and  
 differentiation.

-!- TISSUE SPECIFICITY: Expressed preferentially in hematopoietic  
 progenitors and mature blood cells. Expressed at low levels in the  
 heart, lung, spleen, and thymus and at a higher level in muscle.  
 -!- DEVELOPMENTAL STAGE: Developmentally regulated. Preferential  
 expression in both fetal and adult hematopoietic progenitors and  
 mature blood cells during embryonic and adult hematopoiesis.  
 -!- SIMILARITY: Belongs to the TRIP-Br family.

This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation  
 at the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; AK010535; BAB27012.2;  
 CC DR EMBL; AF322238; AAH31074.1;  
 CC DR EMBL; AF322238; AAH31074.1;  
 CC DR EMBL; AK032980; BAC28109.1;  
 CC DR EMBL; BC012953; AAH12953.1; ALT\_INIT.  
 CC DR MGI; MGI:1919213; Cdca4.  
 CC DR InterPro; IPR009263; SERTA.  
 CC DR Pfam; PF06031; SERTA; 1.  
 CC DR CONFLICT; 196 196; S -> T (in Ref. 3).  
 CC FT CONFLICT; 213 213; T -> A (in Ref. 3).  
 CC SQ SEQUENCE; 237 AA; 26107 MW; 52EC046EFDFF326E CRC64;  
 CC Query Match 13.2%; Score 136; DB 1; Length 237;  
 CC Best Local Similarity 26.1%; Pred. No. 0 01;  
 CC Matches 32; Mismatches 86; Indels 60; Gaps 9;  
 CC保守性 63; 演变性 32;  
 CC 4 GLKRKHSDELEEEERNEWSPA-GLOSYQQALLRISLDKVQRLGRAPSILRRHVLIHTL 62  
 CC 5 GLKRKYSQDQEERGVGGFTVPSYSIQ--RQSLLDMSLVKLQLCHMLVEPNLCRSVLIANTV 62  
 CC QY 63 QQLOQAALR-----LAP----APALPPEPLFLGEEDFSLSATIGSILRELDTSMDGTEPPQ 113  
 CC Db 63 RQIQEEMSQDGVWHGMAPQNVDRAPELADPLQ 119  
 CC QY 114 NPVTPLGLQNEVPPQDP-----VFEALSSRYLGDSQLDDFF 151  
 CC Db 120 NSVSELPVQGSARGORNQSSLWMDSPQENRGSFQKSLLDQIFETLNK--NSSSVEELF 177  
 CC QY 152 LDIDTSAVEKEPAR-----APPEPPHNLFCAPGSWEWNELDHIMEII 193  
 CC Db 178 SDVDSSYYDLDTIVLTGMMMSGTKSSILCNGLEGFAATPPSSTCKS--DLAELDHVVEIL 234  
 CC QY 194 L 194  
 CC Db 235 V 235  
 CC RESULT 12  
 CC AAH55824 PRELIMINARY; PRT; 237 AA.  
 CC AAH55824 PRELIMINARY; PRT; 237 AA.  
 CC AC AAH55824; PRELIMINARY; PRT; 237 AA.  
 CC DT 02-MAR-2004 (TREMBL); 27, Created  
 CC DT 02-MAR-2004 (TREMBL); 27, Last sequence update  
 CC DT 02-MAR-2004 (TREMBL); 27, Last annotation update  
 CC DE Cell division cycle associated 4.  
 CC OS Mus musculus (Mouse).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC OC NCBI\_TaxID=10090;  
 CC RN [1]  
 CC TISSUE=Eye;  
 CC RC SEQUENCE FROM N.A.  
 CC RP MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 CC RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 CC RA Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuller G.D.,  
 CC RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 CC RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 CC RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 CC RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
 CC RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 CC RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 CC RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 CC RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Huiyk S.W.,  
 CC RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 CC RA Fahay J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A.,  
 CC RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 CC RA Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C.,  
 CC RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 CC RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 CC RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 CC RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences";  
 CC RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 CC "-!- FUNCTION: May be involved in molecular regulation of hematopoietic  
 stem cells and progenitor cell lineage commitment and  
 differentiation.

Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Hale L.J., Gibbs R.A., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smailus D.E., Schnurch A., Schein J.E., Jones S.J., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences"; RT and mouse cDNA sequences"; RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RN [2]

RN SEQUENCE FROM N.A.

RP TISSUE=Eye;

RC Strausberg R.; Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.

RL DR EMBL; BC055824; AAH55824.1; -.

KW Cell division.

SQ SEQUENCE 237 AA; 26107 MW; 52EC046EFDFF326E CRC64;

Query Match 13.2%; Score 136; DB 2; Length 237; Best Local Similarity 26.1%; Pred. No. 0.01; Matches 63; Conservative 32; Mismatches 86; Indels 60; Gaps 9;

QY 4 GLKRKHSDLEEEERNEWSPA-GLOSYQQALLRISLDKVQRSLGPRAPSLLRRHVLHNTL 62

Db 5 GLKRKYGQEEGVEGFGTVPSYSLQ--RQSLLDMSLVKLQLCHMLVEPNLCRSVLANTV 62

Qy 63 QQLQAAALR----LAP----APALPPEPLGEEDFSLSATIGSILRELDTSMDGTEPPQ 113

Db 63 RQIQEEMSQDGWVHGNAQPQNNDRAPVERLSTE--ILCRTVRGAAEEHPAPELEDAPLQ 119

Qy 114 NPVTPLGLQNEVPQPDP-----VFLLEALSSRYLGDSGLDDFF 151

Db 120 NSVSELPVGSAFPQRNPQSSLWEMDSSPQENRGSFQRSLDQIFETLENK--NSSSVVELF 177

Qy 152 LDIDTSAVEKEPAR-----APPEPPHNLFCAPGSWEWNELDHIMEII 193

Db 178 SDVDSSYYDLDTVLTGMMSGTKSLLCNGLEGFAAATTPPPSSTCKS---DLAELDHVVEIL 234

Qy 194 L 194

Db 235 V 235

RESULT 13 Q6NXD9 PRELIMINARY; PRT; 361 AA.

AC Q6NXD9; PRELIMINARY; PRT; 361 AA.

DT 05-JUL-2004 (TREMBLrel. 27, Created)

DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)

DE Hypothetical protein sertad2.

GN Name=sertad2;

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysii; Cypriniformes; Cyprinidae; Danio.

OC Cyprinidae; Danio.

OX NCBI\_TaxID=7955;

OX NCBI\_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=22388257; PubMed=1247932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schafer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong Li., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Hale L.J., Gibbs R.A., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Ketteman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S., Krzywinski M.I., Skalska U., Smailus D.E., Schnurch A., Schein J.E., Jones S.J., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences"; RT and mouse cDNA sequences"; RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).

RN [2]

RN SEQUENCE FROM N.A.

RP TISSUE=Kidney;

RC Strausberg R.; Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.

RL DR EMBL; BC067135; AAH67135.1; -.

DR InterPro; IPR009263; SERTA.

DR PFam; PF06031; SERTA; 1.

KW Hypothetical protein.

SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;

Query Match 12.9%; Score 133; DB 2; Length 361; Best Local Similarity 22.3%; Pred. No. 0.029; Matches 81; Conservative 35; Mismatches 71; Indels 176; Gaps 17;

QY 4 GLKRKHSDLEEEERNEWSPA-GLOSYQQALLRISLDKVQRSLGPRAPS 51

Db 5 GAKRK--LDEDEEGLEGKALAAGAGAEGGLSKVSYTLQRQTIFNMSLMKLYNHRAVTEPS 61

QY 52 DRRHVLTHTLQOLOAALR-----L 71

Db 62 LEKRVLINNMLRRIQDELKQEGNLRPLFFPPSPPPDDEVDDESFRQPASFVLSMVAPI 121

QY 72 APAPALP-----PEPL-----FLGEDDFSLSA-----93

Db 122 SQSPALSASSLITSPSSGLSNPAPLEACLTAPPLIEDDNVSLCTSPSPLAPPAPTSRLSPS 181

QY 94 ----TGSILRELD-----TSMDGTEPP-----QNPVPTPL 119

Db 182 VARDSFSSALDEIEELCPSPPLPTATSAGATSPSPLQLCPPSLNSGALDSKDCSKPCSPK 241

QY 120 --GL-----QNEVPQPFD----PVFILE-----ALSSRYLGDSGLDD-FFLDIDTSAVER 161

Db 242 LEGLVLPLAERSAVPNTPETLPPNSLDMSTSPPSASSSGFLTDLADDILFADIDTSMYDF 301

QY 162 EP-----ARAPPEPHNLFCAPGSWEWNELDHIMEII 193

Db 302 DPCTSSSGAAPSKLAPMVTADELLKTFSPYSGAAPAVSN---QPFKMDLTLDHIMEVL 358

RESULT 14 AAH67135 PRELIMINARY; PRT; 361 AA.

AC AAH67135; PRELIMINARY; PRT; 361 AA.

DT 01-JUN-2004 (TREMBLrel. 27, Created)

DT 01-JUN-2004 (TREMBLrel. 27, Last sequence update)

DT 01-JUN-2004 (TREMBLrel. 27, Last annotation update)

DE Hypothetical protein sertad2.

GN Name=sertad2;

OS Brachydanio rerio (Zebrafish) (Danio rerio).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysii; Cypriniformes; Cyprinidae; Danio.

OC Cyprinidae; Danio.

OX NCBI\_TaxID=7955;

OX NCBI\_TaxID=7955;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RX MEDLINE=22388257; PubMed=1247932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shemmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schafer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong Li., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

SEQUENCE FROM N.A.	
RP TISSUE=Kidney;	
RX MEDLINE=22388257; PubMed=12477932;	
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.P., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A., Whiting R.W., Touchman J.W., Green E.D., Dickson M.C., Blakesley R.C., Grimes J., Schmutz J., Myers R.M., Butterfield Y.S., Rodriguez A.C., Young A.C., Shevchenko Y., Bouffard G.G., Krzywinski M.I., Skalska U., Smailus D.E., Schnarch A., Schein J.B., Jones S.J., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences"; Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).	
RT	"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR-2004) to the EMBL/GenBank/DDBJ databases.	
DR EMBL; BC067135; AAH67135.1; -.	
KW Hypothetical protein.	
SQ SEQUENCE 361 AA; 37965 MW; 59CE364B098A43BF CRC64;	
SEQUENCE FROM N.A.	
RC TISSUE=Kidney;	
RA Strausberg R.;	
RL Submitted (MAR	

GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 15, 2004, 14:06:47 ; Search time 142 Seconds  
 (without alignments)  
 488.370 Million cell updates/sec

Title: US-10-069-386-2  
 Perfect score: 1033  
 Sequence: 1 MEGGLKRKHSDLEEEERWE.....APGSWEWNELDHIMEITLGS 196

Scoring table: BLCSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 1568699 seqs, 353819137 residues

Total number of hits satisfying chosen parameters: 1568699.

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
 Listing First 45 summaries

Database : Published Applications AA:  
 1: /cgn2\_6/ptodata/2/pubpaa/us07\_PUBCOMB.pep:  
 2: /cgn2\_6/ptodata/2/pubpaa/pct\_NEW\_PUB.pep:  
 3: /cgn2\_6/ptodata/2/pubpaa/us06\_PUBCOMB.pep:  
 4: /cgn2\_6/ptodata/2/pubpaa/us07\_NEW\_PUB.pep:  
 5: /cgn2\_6/ptodata/2/pubpaa/pctus\_PUBCOMB.pep:  
 6: /cgn2\_6/ptodata/2/pubpaa/us08\_NEW\_PUB.pep:  
 7: /cgn2\_6/ptodata/2/pubpaa/us08\_PUBCOMB.pep:  
 8: /cgn2\_6/ptodata/2/pubpaa/us09A\_PUBCOMB.pep:  
 9: /cgn2\_6/ptodata/2/pubpaa/us09B\_PUBCOMB.pep:  
 10: /cgn2\_6/ptodata/2/pubpaa/us09C\_PUBCOMB.pep:  
 11: /cgn2\_6/ptodata/2/pubpaa/us09\_NEW\_PUB.pep:  
 12: /cgn2\_6/ptodata/2/pubpaa/us10A\_PUBCOMB.pep:  
 13: /cgn2\_6/ptodata/2/pubpaa/us10B\_PUBCOMB.pep:  
 14: /cgn2\_6/ptodata/2/pubpaa/us10C\_PUBCOMB.pep:  
 15: /cgn2\_6/ptodata/2/pubpaa/us10D\_PUBCOMB.pep:  
 16: /cgn2\_6/ptodata/2/pubpaa/us10A\_PUBCOMB.pep:  
 17: /cgn2\_6/ptodata/2/pubpaa/us10\_NEW\_PUB.pep:  
 18: /cgn2\_6/ptodata/2/pubpaa/us11\_NEW\_PUB.pep:  
 19: /cgn2\_6/ptodata/2/pubpaa/us60\_NEW\_PUB.pep:  
 20: /cgn2\_6/ptodata/2/pubpaa/us60\_PUBCOMB.pep:  
 RESULT 1  
 US-10-106-698-5516 ; Sequence 5516, Application US/10106698  
 ; Publication No. US20030109690A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ruben et al.  
 ; TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides  
 ; FILE REFERENCE: PA005P1  
 ; CURRENT APPLICATION NUMBER: US/10/106,698  
 ; CURRENT FILING DATE: 2002-03-27  
 ; PRIOR APPLICATION NUMBER: PCT/US00/26524  
 ; PRIOR FILING DATE: 2000-09-28  
 ; PRIOR APPLICATION NUMBER: US 60/157,137  
 ; PRIOR FILING DATE: 1999-09-29  
 ; PRIOR APPLICATION NUMBER: US 60/163,280  
 ; PRIOR FILING DATE: 1999-11-03  
 ; NUMBER OF SEQ ID NOS: 8564  
 ; SOFTWARE: PatentIn Ver. 3.0  
 ; SEQ ID NO 5516  
 ; LENGTH: 157  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: MISC\_FEATURE  
 ; LOCATION: (55)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: MISC\_FEATURE  
 ; LOCATION: (132)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; US-10-106-698-5516  
 ; QUERY MATCHES:  
 ; Query Match 76.6%; Score 791; DB 14; Length 157;  
 ; Best Local Similarity 98.0%; Pred. No. 1.3e-63;  
 ; Matches 150; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
 ; QY 44 SLGPAPSLRRHVLINHTLQQLQAAIRLAPAPALPPEPLFGEEDFSLSATIGSILRELD 103  
 ; SEQUENCES:  
 ; Sequence 5516, Appl  
 ; Sequence 33709, A  
 ; Sequence 33708, A  
 ; Sequence 568, APP  
 ; Sequence 412, APP  
 ; Sequence 1065, APP  
 ; Sequence 34001, A  
 ; Sequence 2, Appl  
 ; Sequence 31, Appl  
 ; Sequence 627, APP  
 ; Sequence 596, APP  
 ; Sequence 596, APP

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	791	157	14	US-10-106-698-5516	Sequence 5516, Appl
2	645.5	142	14	US-10-029-386-33709	Sequence 33709, A
3	402	111	14	US-10-029-386-33708	Sequence 33708, A
4	216	236	11	US-09-978-360A-568	Sequence 568, APP
5	213	206	9	US-09-731-B72-412	Sequence 412, APP
6	213	236	10	US-09-876-997-412	Sequence 412, APP
7	213	278	15	US-10-296-115-1065	Sequence 1065, APP
8	205	198	14	US-10-029-386-34001	Sequence 34001, A
9	136	237	13	US-10-076-069-2	Sequence 2, Appl
10	125	121	14	US-10-170-385-31	Sequence 31, Appl
11	125	121	16	US-10-755-889-627	Sequence 627, APP
12	106	103	254	9	US-09-925-302-596
13	106	103	10	US-09-925-302-596	Sequence 596, APP

OTHER INFORMATION: MAP TO AC010271.5  
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.64  
 US-10-029-386-33708

Query Match 38.9%; Score 402; DB 14; Length 111;  
 Best Local Similarity 89.9%; Pred. No. 1.4e-28;  
 Matches 80; Conservative 0; Mismatches 9; Indels 0; Gaps 0

Qy 1 MEGGLKRKHSDLEEEERWEMSPAGLQSYQQALLRISLDKVQRSLGRAPSURHVLHN 60  
 Db 3 MVGGGLKRKHSDLEEEERWEMSPAGLQSYQQALLRISLDKVQRSLGRAPSURHVLHN 62

Query Match 38.9%; Score 402; DB 14; Length 111;  
 Best Local Similarity 89.9%; Pred. No. 1.4e-28;  
 Matches 80; Conservative 0; Mismatches 9; Indels 0; Gaps 0

Qy 1 MEGGLKRKHSDLEEEERWEMSPAGLQSYQQALLRISLDKVQRSLGRAPSURHVLHN 60  
 Db 3 MVGGGLKRKHSDLEEEERWEMSPAGLQSYQQALLRISLDKVQRSLGRAPSURHVLHN 62

RESULT 2  
 JS-10-029-386-33709  
 Sequence 33709, Application US/10029386  
 Publication No. US20030194704A1

GENERAL INFORMATION:  
 APPLICANT: Penn, Sharron G.  
 APPLICANT: Rank, David R.  
 APPLICANT: Hanzel, David K.  
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GENE EXPRESSION ANALYSIS TWO

FILE REFERENCE: AEOMICA-X-2  
 CURRENT APPLICATION NUMBER: US/10/029,386  
 CURRENT FILING DATE: 2001-12-20  
 NUMBER OF SEQ ID NOS: 34288  
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 SEQ ID NO: 33709  
 LENGTH: 142  
 TYPE: PRT  
 ORGANISM: Homo sapiens

FEATURE:  
 OTHER INFORMATION: MAP TO AC010271.5  
 OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.64  
 JS-10-029-386-33709

Query Match 62.5%; Score 645.5; DB 14; Length 142;  
 Best Local Similarity 96.8%; Pred. No. 1.7e-50;  
 Matches 122; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 71 LAPAPALPPEPLFLGEEDFSLSATIGSILRELDTSMDGTEPPQNVPVTPLGLQNEVPPQPD 130  
 Db 18 LPPPLPC-PPEPLFLGEEDFSLSATIGSILRELDTSMDGTEPPQNVPVTPLGLQNEVPPQPD 76

Qy 131 PVFLEALSSRYLGDSGLDDFFLDIDTSAVEKEPARAPPEPHNLFCAPGSWEWNELDHIM 190  
 Db 77 PVFLEALSSRYLGDSGLDDFFLDIDTSAVEKEPARAPPEPHNLFCAPGSWEWNELDHIM 136

Qy 191 EIIIGS 196  
 Db 137 EIIIGS 142

Remaining Prior Application data removed - See File Wrapper or PALM.  
 NUMBER OF SEQ ID NOS: 810  
 SOFTWARE: Patent .pm  
 SEQ ID NO 568  
 LENGTH: 236  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: SIGNAL  
 LOCATION: -31...-1  
 FEATURE:  
 NAME/KEY: misc\_feature  
 LOCATION: (28, 30, 40, 67, 86, 117, 120)  
 OTHER INFORMATION: unknown  
 US-09-978-360A-568

RESULT 3  
 JS-10-029-386-33708  
 Sequence 33708, Application US/10029386  
 Publication No. US20030194704A1

GENERAL INFORMATION:  
 APPLICANT: Penn, Sharron G.  
 APPLICANT: Rank, David R.  
 APPLICANT: Hanzel, David K.  
 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR GENE EXPRESSION ANALYSIS TWO

FILE REFERENCE: AEOMICA-X-2  
 CURRENT APPLICATION NUMBER: US/10/029,386  
 CURRENT FILING DATE: 2001-12-20  
 NUMBER OF SEQ ID NOS: 34288  
 SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
 SEQ ID NO 33708  
 LENGTH: 111  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:

Query Match 20.9%; Score 216; DB 11; Length 236;  
 Best Local Similarity 32.6%; Pred. No. 2.4e-11;  
 Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;

Qy 4 GLKRKHSDEEE---ERWEWSAGLQSYQQA---LLRISLDKVRQSLGPRAPI-----

5 GLRKREEEEEKEPLAVDSW-WLDPGHAAVAQAPPVASSSLFDLSVLKHHSLQXSXPD 63 ; PRIORITY FILING DATE: 2000-12-07  
 52 LRRHVLIHNTLQQLQAALRLLAPAPALPPEL-----FLEEDFSLSATIGSILRE 101 ; PRIORITY APPLICATION NUMBER: US 60/187,470  
 64 LRHLVXNTLRIQAS--MAPAAALPPVPTPPAAXVADNLASSDAALSASMAXLLED 121 ; PRIORITY FILING DATE: 2000-03-06  
 ; PRIORITY APPLICATION NUMBER: US 60/169,629  
 ; NUMBER OF SEQ ID NOS: 482  
 ; SOFTWARE: Patent .pm  
 ; SEQ ID NO: 412  
 ; LENGTH: 236  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-876-997-412

Query Match 20.6%; Score 213; DB 10; Length 236;  
 Best Local Similarity 32.6%; Pred. No. 4.6e-11;  
 Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;

Qy 4 GLRKHSDLEEEE---ERWEWSPLAGLQSYQQA-----LRLRISLDKVQRSLGPRAPS 51  
 5 GLRKREEEEEKEPLAVDSW-WLDPGHTAVAQAPPVASSSSLFDLSVLKHHSLQQSEPD 63  
 Db

Qy 52 LRRHVLIHNTLQQLQAALRLLAPAPALPPEL-----FLGEEDFSLSATIGSILRE 101  
 64 LRHLVXNTLRIQAS--MAPAAALPPVPTPPAAXVADNLASSDAALSASMAXLLED 121  
 ; FILE REFERENCE: 78.US3.REG  
 ; CURRENT APPLICATION NUMBER: US/09/731,872  
 ; CURRENT FILING DATE: 2000-12-07  
 ; PRIOR APPLICATION NUMBER: US 60/169,629  
 ; PRIOR FILING DATE: 1999-12-08  
 ; PRIOR APPLICATION NUMBER: US 60/187,470  
 ; PRIOR FILING DATE: 2000-03-06  
 ; NUMBER OF SEQ ID NOS: 482  
 ; SOFTWARE: Patent .pm

RESULT 5  
 US-09-731-872-412 ; Sequence 412, Application US/09731872  
 ; Patent No. US20020102604A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, Jean Baptiste  
 ; APPLICANT: Bougueret, Lydie  
 ; APPLICANT: Jobert, Severin  
 ; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS  
 ; FILE REFERENCE: 78.US3.REG  
 ; CURRENT APPLICATION NUMBER: US/09/731,872  
 ; CURRENT FILING DATE: 2000-12-07  
 ; PRIOR APPLICATION NUMBER: US 60/169,629  
 ; PRIOR FILING DATE: 1999-12-08  
 ; PRIOR APPLICATION NUMBER: US 60/187,470  
 ; PRIOR FILING DATE: 2000-03-06  
 ; NUMBER OF SEQ ID NOS: 482  
 ; SEQ ID NO: 412  
 ; LENGTH: 236  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-09-731-872-412

Query Match 20.6%; Score 213; DB 9; Length 236;  
 Best Local Similarity 32.6%; Pred. No. 4.6e-11;  
 Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;

Qy 4 GLRKHSDLEEEE---ERWEWSPLAGLQSYQQA-----LRLRISLDKVQRSLGPRAPS 51  
 5 GLRKREEEEEKEPLAVDSW-WLDPGHTAVAQAPPVASSSSLFDLSVLKHHSLQQSEPD 63  
 Db

Qy 52 LRRHVLIHNTLQQLQAALRLLAPAPALPPEL-----FLGEEDFSLSATIGSILRE 101  
 64 LRHLVXNTLRIQAS--MAPAAALPPVPTPPAAXVADNLASSDAALSASMAXLLED 121  
 ; FILE REFERENCE: 78.US3.REG  
 ; CURRENT APPLICATION NUMBER: US/09/731,872  
 ; CURRENT FILING DATE: 2000-12-07  
 ; PRIOR APPLICATION NUMBER: US 60/169,629  
 ; PRIOR FILING DATE: 1999-12-08  
 ; PRIOR APPLICATION NUMBER: US 60/187,470  
 ; PRIOR FILING DATE: 2000-03-06  
 ; NUMBER OF SEQ ID NOS: 482  
 ; SOFTWARE: Patent .pm

RESULT 7  
 US-10-296-115-1065 ; Sequence 1065, Application US/10296115  
 ; Publication No. US20040053248A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: HYSEQ INC  
 ; TITLE OF INVENTION: NO. US20040053248A1el Nucleic Acids and Polypeptides  
 ; FILE REFERENCE: 784.PCT  
 ; CURRENT APPLICATION NUMBER: US/10/296,115  
 ; CURRENT FILING DATE: 2002-11-18  
 ; PRIOR APPLICATION NUMBER: US09/488,725  
 ; PRIOR FILING DATE: 2000-01-21  
 ; PRIOR APPLICATION NUMBER: US09/552,317  
 ; PRIOR FILING DATE: 2000-04-25  
 ; NUMBER OF SEQ ID NOS: 1478  
 ; SEQ ID NO: 1065  
 ; LENGTH: 278  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; US-10-296-115-1065

Query Match 20.6%; Score 213; DB 15; Length 278;  
 Best Local Similarity 32.6%; Pred. No. 5.6e-11;  
 Matches 76; Conservative 36; Mismatches 69; Indels 52; Gaps 13;

Qy 4 GLRKHSDLEEEE---ERWEWSPLAGLQSYQQA-----LRLRISLDKVQRSLGPRAPS 51  
 52 LRRHVLIHNTLQQLQAALRLLAPAPALPPEL-----FLGEEDFSLSATIGSILRE 101  
 64 LRHLVXNTLRIQAS--MAPAAALPPVPTPPAAXVADNLASSDAALSASMAXLLED 163  
 Db

Qy 102 LDTSMGDTEPPQNPNVTPLGLQNEVPP-----QPDPVFLEAL--SSRYLGDSGLDDFFL 152  
 122 L-SHIEGLSQAPQP----LADEGPPGRSIGGAAPSLGALDLGPATGCLLDGLEGLFE 175  
 ; FILE REFERENCE: 78.US4.CIP  
 ; CURRENT APPLICATION NUMBER: US/09/876,997  
 ; CURRENT FILING DATE: 2001-06-08  
 ; PRIOR APPLICATION NUMBER: US 09/731,872

Qy 153 DIDTSAVEKE---PARAPPEP-PHNLFCAPGSWE----WNELDHIMEIIILGS 196  
 176 DIDTSMYDNELWAPASEGLKPGPED--GPGKEEAPELDEAELDYLMDVLVGT 225  
 ; FILE REFERENCE: 78.US3.REG  
 ; CURRENT APPLICATION NUMBER: US/09/731,872  
 ; CURRENT FILING DATE: 2000-12-07  
 ; PRIOR APPLICATION NUMBER: US 60/169,629  
 ; PRIOR FILING DATE: 1999-12-08  
 ; NUMBER OF SEQ ID NOS: 482  
 ; SOFTWARE: Patent .pm

RESULT 6  
 US-09-876-997-412 ; Sequence 412, Application US/09876997  
 ; Publication No. US20030152921A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Dumas Milne Edwards, Jean Baptiste  
 ; APPLICANT: Bougueret, Lydie  
 ; APPLICANT: Jobert, Severin  
 ; TITLE OF INVENTION: FULL-LENGTH HUMAN CDNAS ENCODING POTENTIALLY SECRETED PROTEINS  
 ; FILE REFERENCE: 78.US4.CIP  
 ; CURRENT APPLICATION NUMBER: US/09/876,997  
 ; CURRENT FILING DATE: 2001-06-08  
 ; PRIOR APPLICATION NUMBER: US 09/731,872

Qy 47 GLRKREEEEEKEPLAVDSW-WLDPGHAAVAQAPPVASSSSLFDLSVLKHHSLQQSEPD 105  
 52 LRRHVLIHNTLQQLQAALRLLAPAPALPPEL-----FLGEEDFSLSATIGSILRE 101  
 64 LRHLVXNTLRIQAS--MAPAAALPPVPTPPAAXVADNLASSDAALSASMAXLLED 163  
 Db

Qy 102 LDTSMGDTEPPQNPNVTPLGLQNEVPP-----QPDPVFLEAL--SSRYLGDSGLDDFFL 152  
 164 L-SHIEGLSQAPQP----LADEGPPGRSIGGAAPSLGALDLGPATGCLLDGLEGLFE 217  
 ; FILE REFERENCE: 78.US4.CIP  
 ; CURRENT APPLICATION NUMBER: US/09/876,997  
 ; CURRENT FILING DATE: 2001-06-08  
 ; PRIOR APPLICATION NUMBER: US 09/731,872

Qy 153 DIDTSAVEKE---PARAPPEP-PHNLFCAPGSWE----WNELDHIMEIIILGS 196  
 ; FILE REFERENCE: 78.US3.REG  
 ; CURRENT APPLICATION NUMBER: US/09/731,872  
 ; CURRENT FILING DATE: 2000-12-07  
 ; PRIOR APPLICATION NUMBER: US 60/169,629  
 ; PRIOR FILING DATE: 1999-12-08  
 ; NUMBER OF SEQ ID NOS: 482  
 ; SOFTWARE: Patent .pm



Qy 146 GLDD-FFLDIDTSAVEKEP-----ARAPPEPHNLF-----CAPGSNEW 183  
Db 242 TLDDILFADIDTSMYDFDPCTSSSGTASRMAPVSADDLKLTLAPYSSQPVTTPSQQPFRMDL 301

Qy 184 NELDHIMEIILGS 196  
Db 302 TELDHIMEVVLGS 314

RESULT 11  
Sequence 627, Application US/10755889  
Publication No. US20040171823A1

GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES ASSOCIATED WITH THE NF-kB  
FILE REFERENCE: D0284 NP  
CURRENT APPLICATION NUMBER: US/10/755, 889  
CURRENT FILING DATE: 2004-01-13  
PRIOR APPLICATION NUMBER: U.S. 60/440, 068  
PRIOR FILING DATE: 2003-01-14  
PRIOR APPLICATION NUMBER: U.S. 60/469, 757  
PRIOR FILING DATE: 2003-05-12  
NUMBER OF SEQ ID NOS: 823  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 627  
LENGTH: 314  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-755-889-627

Query Match Score 106; DB 9; Length 254;  
Best Local Similarity 25.6%; Pred. No. 0.24;  
Matches 65; Conservative 32; Mismatches 75; Indels 82; Gaps 13;

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-596

Qy 4 GLKRKHSDELLEEEERNEWSPAGLQ---SY---QQALLRISLDKVQRSLGPRAPSILRRHVL 57  
Db 18 GLRKVCVGHEEDV---EGALAGLKTVSSYSLQRQSLLDMSLVKLQLCHMLVEPNLCRSVL 74

Qy 58 IANTLQQLQAAL-----RLAP--APALPPEPLFLGEEDFLSATIGSILRELDTSMDG 108  
Db 75 IANTVHQIQQEEMTQDGQTWRTRVAPQAAERAPXDRLVSTE-----ILCRAAWGQEG 123

Qy 109 TEPE-----PQNPT-----PLGLQNEVPPQQDP-----VFEALAS 138  
Db 124 AHPAPGLGDGHTQGPvSDLCFVTSQAQAPRHLOSSAWEMDGPRENRGSFHKSLLDQIFETLE 183

Qy 139 SRYLGDSGLDDFFFLDIDTSAVEKEP-----AR-----APPEPPHNLFCAPGS 180  
Db 184 TK--NPSCMEEELFSVDSPYYDLDTVLTMGGARPGPCBEGLEGЛАPATPGPSSSCKS-- 239

Query Match Score 125; DB 16; Length 314;  
Best Local Similarity 22.7%; Pred. No. 0.006;  
Matches 71; Conservative 34; Mismatches 88; Indels 120; Gaps 12;

Qy 2 EGGLKRKHSDELLEEEERNEWSPAGLQS-----YQOALLRISLDKVQRSLGPRAPSILRRH 55  
Db 4 KGG-KRKFDDEHEDGLEGRKIVSPCDGPSKVSYTLYRQQTIFNISLMKLYNHRPLTEPSLQKT 62

RESULT 13  
US-09-925-302-596

Qy 56 VLIANTLQQLQAALR-----  
Db 63 VLIINMRLRIQEELKQEGSLRPMETPSSQPTEPSSYREAPPAFSHLASPSSHPCDLGS 122

Qy 71 -----LAPA-----PALPPE--PLFLGEEDFSLSATIGSILRELD 103  
Db 123 TTPLEACLT PASLLEDDDDTFCTSQAQM QPTAPT KLSPPALLEPKD-SFSSALDEIEELCP 181

Qy 104 TSMD-----GTEPPQNPNPVTPLGLQNEVPPQ----PDPVFLAE-----SSRYLGDS 145  
Db 182 TSTSTAATAATDSVKGTSSSEAGTQKLDPQESRADDSKLMDSLPGNFELITSTGFLLTD 241

Qy 146 GLDD-FFLDIDTSAVEKEP-----ARAPPEPHNLF-----CAPGSNEW 183  
Db 242 TLDDILFADIDTSMYDFDPCTSSSGTASRMAPVSADDLKLTLAPYSSQPVTTPSQQPFRMDL 301

Qy 184 NELDHIMEIILGS 196  
Db 302 TELDHIMEVVLGS 314

RESULT 12  
US-09-925-302-596  
Sequence 596, Application US/09925302  
Patent No. US20020044941A1

GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA104  
CURRENT APPLICATION NUMBER: US/09/925, 302  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05918  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124, 270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 896  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 596  
LENGTH: 254  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (105)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-302-596

Qy 4 GLKRKHSDELLEEEERNEWSPAGLQ---SY---QQALLRISLDKVQRSLGPRAPSILRRHVL 57  
Db 18 GLRKVCVGHEEDV---EGALAGLKTVSSYSLQRQSLLDMSLVKLQLCHMLVEPNLCRSVL 74

QY 58 IHNTLQLQQAAL---RLAP--APALPPEPLFLGEDEFSLSATIGSILRELDTSMDG 108  
 DB 75 IANTVHQIQQEMTQDGTVRTPAQAAERAPXDRLVSTE-----ILCRAAWGQEG 123  
 QY 109 TEP----PQNPTV-----PLGLQNEVPQPQDP-----VFELEALS 138  
 DB 124 AHPAPGLGDGHTQGPVSDLCPVTSAQAPRHLQSSAWEMDGPRENRSFHKSLDQIFETLE 183  
 QY 139 SRYLGDSGLDDFFLIDITSAVEKEP-----AR-----APPEPPHNLFCAPGS 180  
 DB 184 TK--NPSCMEEELFSDVDSPYYDLDTVLTCMMGGARPQCEGLEGLAPATPGSSCKS-- 239  
 QY 181 WENWLDHIMEIIL 194  
 DB 240 -DLGELDHVVEILV 252

RESULT 14  
 US-10-106-698-4456  
 Sequence 4456, Application US/10106698  
 Publication No. US20030109690A1  
 GENERAL INFORMATION:  
 APPLICANT: Ruben et al.  
 TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptide  
 FILE REFERENCE: PA005P1  
 CURRENT APPLICATION NUMBER: US/10/106,698  
 CURRENT FILING DATE: 2002-03-27  
 PRIOR APPLICATION NUMBER: PCT/US00/26524  
 PRIOR FILING DATE: 2000-09-28  
 PRIOR APPLICATION NUMBER: US 60/157,137  
 PRIOR FILING DATE: 1999-09-29  
 PRIOR APPLICATION NUMBER: US 60/163,280  
 PRIOR FILING DATE: 1999-11-03  
 NUMBER OF SEQ ID NOS: 8564  
 SOFTWARE: Patentin Ver. 3.0  
 SEQ ID NO 4456  
 LENGTH: 254  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE: MISC\_FEATURE  
 LOCATION: (105)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 US-10-106-698-4456

Query Match 10.3%; Score 106; DB 14; Length 254;  
 Best Local Similarity 25.6%; Pred. No. 0.24;  
 Matches 65; Conservative 32; Mismatches 75; Indels 82; Gaps 13;  
 QY 4 GLKRKHSDSLLEEEERWEWSPEAGLQ---SY---QQALLRISLDKVQRSLGPRAPSLLRRHVL 57  
 DB 18 GLKRKCVGHEEDV---EGALAGLKTVSSYSLQRQSLLDMSLVLKQLCHMLVEPNLCRSVL 74  
 QY 58 IHNTLQLQQAAL-----RLAP--APALPPEPLFLGEDEFSLSATIGSILRELDTSMDG 108  
 DB 75 IANTVHQIQQEMTQDGTVRTPAQAAERAPXDRLVSTE-----ILCRAAWGQEG 123  
 QY 109 TEP----PQNPTV-----PLGLQNEVPQPQDP-----VFELEALS 138  
 DB 124 AHPAPGLGDGHTQGPVSDLCPVTSAQAPRHLQSSAWEMDGPRENRSFHKSLDQIFETLE 183  
 QY 139 SRYLGDSGLDDFFLIDITSAVEKEP-----AR-----APPEPPHNLFCAPGS 180  
 DB 184 TK--NPSCMEEELFSDVDSPYYDLDTVLTCMMGGARPQCEGLEGLAPATPGSSCKS-- 239  
 QY 181 WENWLDHIMEIIL 194  
 DB 240 -DLGELDHVVEILV 252

Search completed: November 15, 2004, 14:19:09  
 Job time : 147 secs

Query Match 10.2%; Score 105; DB 13; Length 241;  
 Best Local Similarity 25.9%; Pred. No. 0.28;  
 Matches 65; Conservative 31; Mismatches 79; Indels 76; Gaps 13;  
 QY 4 GLKRKHSDSLLEEEERWEWSPEAGLQ---SY---QQALLRISLDKVQRSLGPRAPSLLRRHVL 57  
 DB 5 GLKRKCVGHEEDV---EGALAGLKTVSSYSLQRQSLLDMSLVLKQLCHMLVEPNLCRSVL 61  
 QY 58 IHNTLQLQQAAL-----RLAP--APALPPEPLFLGEDEFSLSATIGSILRELDTSMDG 108  
 DB 62 IANTVHQIQQEMTQDGTVRTPAQAAERAPL-----DRLVST---EILCRAAWGQEGAHP 113  
 QY 112 -----PQNPTV-----PLGLQNEVPQPQDP-----VFELEALSSRY 141  
 DB 114 APGLGDGHTQGPVSDLCPVTSAQAPRHLQSSAWEMDGPRENRSFHKSLDQIFETLET 172  
 QY 142 LGDSGLDDFFLIDITSAVEKEP-----AR-----APPEPHNLFCAPGSWEW 183  
 DB 173 -NPSCMEEELFSDVDSPYYDLDTVLTCMMGGARPQCEGLEGLAPATPGSSCKS--DL 228  
 QY 184 NELDHIMEIIL 194  
 DB 229 GELDHVVEILV 239

RESULT 15  
 US-10-076-069-4  
 Sequence 4, Application US/10076069

GenCore version 5.1.6  
 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model  
 Run on: November 15, 2004, 13:58:11 ; Search time 40 Seconds  
 (without alignments)  
 324.958 Million cell updates/sec

Title: US-10-069-386-2  
 Perfect score: 1033  
 Sequence: 1 MEGGLKRKHSDLEEEERWE.....APGSWENNLDHIMEITLGS 196  
 Scoring table: BLOSUM62  
 Gapext 0.5  
 Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
 Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing First 45 summaries

Database : Issued Patents AA:  
 1: /cgn2-6/ptodata/1/iaa/5A-COMB.pep:  
 2: /cgn2-6/ptodata/1/iaa/5B-COMB.pep:  
 3: /cgn2-6/ptodata/1/iaa/6A-COMB.pep:  
 4: /cgn2-6/ptodata/1/iaa/6B-COMB.pep:  
 5: /cgn2-6/ptodata/1/iaa/PCTUS-COMB.pep:  
 6: /cgn2-6/ptodata/1/iaa/backfiles1.pep:  
 \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94	9.1	433	4 US-08-259-451-3	Sequence 3, Appli
2	93	9.0	433	1 US-07-672-483-4	Sequence 4, Appli
3	93	9.0	2618	3 US-09-413-814-28	Sequence 28, Appli
4	92.5	9.0	432	5 PCT-US95-04910-13	Sequence 13, Appli
5	90.5	8.8	370	4 US-09-377-285B-16	Sequence 16, Appli
6	90.5	8.8	947	2 US-08-887-518-2	Sequence 2, Appli
7	90.5	8.8	947	2 US-09-023-321-2	Sequence 2, Appli
8	90.5	8.8	947	2 US-09-032-475-2	Sequence 2, Appli
9	90.5	8.8	947	3 US-09-257-703-1	Sequence 1, Appli
10	90.5	8.8	947	4 US-09-871-889A-1	Sequence 1, Appli
11	89.5	8.7	215	4 US-08-778-717-9	Sequence 2, Appli
12	88.5	8.6	897	4 US-09-849-602-18	Sequence 18, Appli
13	88	8.5	1130	4 US-09-976-594-280	Sequence 280, Appli
14	87.5	8.5	709	4 US-10-118-328-4	Sequence 4, Appli
15	86.5	8.4	136	4 US-08-259-451-5	Sequence 5, Appli
16	84	8.1	337	4 US-09-543-681A-744	Sequence 7444, Appli
17	84	8.1	514	4 US-09-252-991A-25281	Sequence 25281, Appli
18	84	8.1	628	4 US-09-345-473E-48	Sequence 48, Appli
19	84	8.1	2152	3 US-09-036-987A-3	Sequence 3, Appli
20	84	8.1	2152	3 US-09-370-700-3	Sequence 3, Appli
21	84	8.1	2152	4 US-09-603-207-3	Sequence 3, Appli
22	83	8.0	969	2 US-08-548-159-1	Sequence 1, Appli
23	83	8.0	986	2 US-08-548-159-3	Sequence 3, Appli
24	83	8.0	1012	3 US-08-811-481-16	Sequence 16, Appli
25	83	8.0	1012	4 US-09-876-527-16	Sequence 16, Appli
26	82.5	8.0	1130	4 US-09-538-092-1142	Sequence 834, Appli
27	82.5	8.0	2142	4 US-09-538-092-1142	Sequence 1142, Appli

#### ALIGNMENTS

RESULT 1  
 US-08-259-451-3  
 ; Sequence 3, Application US/08259451  
 ; Patent No. 6406841  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Lee, Helen H.  
 ; APPLICANT: Swanson, Priscilla A.  
 ; APPLICANT: Idler, Kenneth B.  
 ; APPLICANT: Rosenblatt, Joseph D.  
 ; APPLICANT: Chen, Irvin S. Y.  
 ; APPLICANT: Golde, David W.  
 ; APPLICANT: Robertson, Eugene F.  
 ; APPLICANT: Stephens, John E.  
 ; APPLICANT: Chan, Emerson W.  
 ; APPLICANT: Buytendorp, Mark H.  
 ; APPLICANT: Johnson, Joan E.  
 ; APPLICANT: Motley, Cheryl T.  
 ; APPLICANT: Peterson, Bryan  
 ; APPLICANT: Edwards, Michelle  
 ; APPLICANT: Guidinger, Peggy  
 ; APPLICANT: Tate, Cynthia  
 ; TITLE OF INVENTION: HTLV-IINRA Compositions  
 ; TITLE OF INVENTION: and Assays for Detecting HTLV Infection  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Abbott Laboratories  
 ; STREET: One Abbott Park Road  
 ; CITY: Abbott Park  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60064  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
 ; COMPUTER: IBM  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: WordPerfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/259, 451  
 ; FILING DATE: 20-JUL-1994  
 ; CLASSIFICATION: 436  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/086, 415  
 ; FILING DATE: 01-JUL-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Daniel W. Collins  
 ; REGISTRATION NUMBER: 31, 912  
 ; REFERENCE/DOCKET NUMBER: 5381.US.P1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (708) 937-6365



APPLICANT: States of America as represented by the Secretary, Department of Health and Human Services

TITLE OF INVENTION: ISOLATION AND CHARACTERIZATION OF A NOVEL PRIMATE T-CELL LYMPHOTROPIC VIRUS AND THE USE OF THIS VIRUS OR COMPONENTS THEREOF IN DIAGNOSTIC ASSAYS

TITLE OF INVENTION: NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORGAN & FINNEGAN

STREET: 345 PARK AVENUE

CITY: NEW YORK

STATE: NEW YORK

COUNTRY: USA

ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: WORDPERFECT 5.1

APPLICATION NUMBER: PCT/US95/04910

FILING DATE: 21-APR-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US08/231,526

FILING DATE: 22-APR-1994

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: WILLIAM S. FEILER

REGISTRATION NUMBER: 26,728

REFERENCE/DOCKET NUMBER: 2026-4125PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 751-6849

TELEX: 421792

INFORMATION FOR SEQ ID NO: 13:

SEQUENCE CHARACTERISTICS:

LENGTH: 432 amino acids

TYPE: amino acids

STRANDEDNESS: unknown

TOPOLOGY: unknown

PCT-US95-04910-13

Query Match 9.0%; Score 92.5; DB 5; Length 432;

Best Local Similarity 21.5%; Pred. No. 0.92; Matches 40; Conservative 19; Mismatches 48; Indels 79; Gaps 7; Gaps 7;

RESULT 6  
US-08-887-518-2

; Sequence 2, Application US/08887518  
; Patent No. 5843721  
; GENERAL INFORMATION:  
; APPLICANT: Rothe, Mike  
; APPLICANT: Wu, Lin

; TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
; STREET: 268 BUSH STREET, SUITE 3200  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: USA  
; ZIP: 94104

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:  
; FILING DATE:  
; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:  
; NAME: OSMAN, RICHARD A  
; REGISTRATION NUMBER: 36,627

RESULT 5  
US-09-377-285B-16  
; Sequence 16, Application US/09377285B  
; Patent No. 6720175  
; GENERAL INFORMATION:  
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE  
; APPLICANT: WORLEY, Paul  
; APPLICANT: TU, Jian

APPLICANT: XIAO, Bo  
APPLICANT: LEAHY, Daniel  
APPLICANT: BENKEN, Jutta  
APPLICANT: LANAHAN, Anthony  
TITLE OF INVENTION: NUCLEAR ACID MOLECULE ENCODING HOMER 1b PROTEIN (AS AMENDED)  
FILE REFERENCE: JHU1580-4  
CURRENT APPLICATION NUMBER: US/09/377,285B  
CURRENT FILING DATE: 1999-08-18  
PRIOR APPLICATION NUMBER: US 60/138,426  
PRIOR APPLICATION NUMBER: US 60/138,493  
PRIOR FILING DATE: 1999-06-10  
PRIOR APPLICATION NUMBER: US 60/138,494  
PRIOR FILING DATE: 1999-06-10  
PRIOR APPLICATION NUMBER: US 60/097,334  
PRIOR FILING DATE: 1998-08-18  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 16  
LENGTH: 370  
TYPE: PRT  
ORGANISM: Rattus norvegicus  
US-09-377-285B-16

Query Match 8.8%; Score 90.5; DB 4; Length 370;  
Best Local Similarity 23.5%; Pred. No. 1.2;  
Matches 43; Conservative 29; Mismatches 62; Indels 49; Gaps 8;

Qy 33 LIRISLDKV-QRSLGGRAPSLLRRHVLIIHQALRLLAPAPLPP----EPLFLG 85  
Db 102 MVNMHLKVARREIGTLA-----TVVRLPPSQKVIPPESSLPPLTPTYCRKPL--- 147

Qy 86 EEDFLSLSATIGSILRELDTSMDGT----EPPQNPNPVTPGLQNEVPPQ-PDPVFLEAL-- 137  
Db 148 -NFACFLDDVGHGKVKDLSLSTSRTGTLRKSIIKAPATPASATLGRPRPIEPVQLPAVFD 205

Qy 138 -----SRYRLGDGGLDDFFLDIDTSAVEKEPARAPPEP----PHNLFCAPG 179  
Db 206 GKLSAASSSSLASAGSAEAGSGIPOSXKGQVAPATPPPPIAPVTPPPPLAEIFULLPP 265

Qy 180 SWE 182  
Db 266 PME 268

Query Match 8.8%; Score 90.5; DB 4; Length 370;  
Best Local Similarity 23.5%; Pred. No. 1.2;  
Matches 43; Conservative 29; Mismatches 62; Indels 49; Gaps 8;

Qy 47 PRAPSLRRHVLIIHQALRLLAPAP----LPPEPLFLGEEDFSLSA 93  
Db 13 FKAP---RGLSTHHLNFLQASYRQLQGPSPDFDFQQLRKTPIWNPIDYSILLA 69

Qy 94 TI-----GSILRELDTSMDGTTEPPQNPTVTPGLQNEVPPQDPVFLEALSSRYLGD 144  
Db 70 SLIPKGYPGRTSEIINVLRINQASPTPPAPSPL-----EPANPPPL--- 111

Qy 145 SGLDDFFLDIDTSAVEKEPARAPPEP----PHNLFCAPGSEWNE 185  
Db 112 -----QQPS-APPEPHTPPPYIEPPATHCLPILPHGAPSAAHRPWQMD 154

Qy 186 LDHIME 191  
Db 155 LQAIKQ 160

REFERENCE/DOCKET NUMBER: T97-008  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 343-4341  
TELEFAX: (415) 343-4342  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 947 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
8-887-518-2

STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: peptide  
 US-09-023-321-2

	Query	Match	Score	DB	Length
Qy	3 GGLKRKHISDLEEEERNEWSPAGLQSYQQALLRISLDKXVQRSLRAPSLLGPRASLRRHVLJHNTL	8.8% ; Best Local Similarity 25.8% ; Matches 54 ;	90.5 ; Pred. No. 4 ; Conservative 17 ;	DB 2 ;	Length 947 ;
Db	659 GGLK---SPWRGEYKEPRHPPPNQANYHQ-----TLHAQPRLSPRAPGPRPAEETTGRA				62
Qy	63 QQLQAAILRLLAPAPALPPE-----PLFLGEEDDFSLSATIGSILRELDTSMDGTEPPQNPNVT				117
Db	711 PKLQ-----PPLPPEPKNSPPLTLSKEE-----SGMWEPLPLSSLEPAPRNPS				758

3 GGLRKHSDEEEERWEWSSPAGLQSQQQALLRISLDKVQRSLGPRAPSLLRRHVLIHNTL 62  
 559 GGLK---SPWRGEYKEPRHPPPNQANYHQ----TLHAQPRELSPRAPGPRPAEETTGRA 710  
 63 QQLQAALRLAPAPALPPE----PLFLGEEDFSLSATIGSILRELDTSMDGTEPPQNTPVT 117  
 711 PRLQ----PPLPPEPKNSPPLTLSKEE----SGMWEPLPLSLEPAPARNFSS 758  
 118 PLGLQNEVPPQ----PDPPVFILEALSRY----LGDSGLDDFFLDIDTSAVEKEPA 164  
 759 P-ERKATVPEQELQOLEIEFLNSLQQPSLEEQQIILSCLSIDSLSSWNS 843  
 165 RAPPEPHNLFCAPGSW----EWN 184  
 815 KASQSSRDTLSSGVHSWSSQEARSSSWN 843  
 RESULT 8  
 US-09-032-475-2  
 ; Sequence 2, Application US/09032475  
 ; Patent No. 5854003  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rothe, Mike  
 ; APPLICANT: Wu, Lin  
 ; TITLE OF INVENTION: NIK Proteins, Nucleic Acids and Methods  
 ; NUMBER OF SEQUENCES: 4  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: SCIENCE & TECHNOLOGY LAW GROUP  
 ; STREET: 268 BUSH STREET, SUITE 3200  
 ; CITY: SAN FRANCISCO  
 ; STATE: CALIFORNIA  
 ; COUNTRY: USA  
 ; ZIP: 94104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/09/032,475  
 ; FILING DATE:  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: 08/887,518  
 ; FILING DATE:  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: OSMAN, RICHARD A  
 ; REGISTRATION NUMBER: 36,627  
 ; REFERENCE/DOCKET NUMBER: T97-008  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (415) 343-4341  
 ; TELEFAX: (415) 343-4342  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 947 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: peptide  
 ; US-09-032-475-2  
 ; Query Match Score 90.5;  
 ; Best Local Similarity 25.8%;  
 ; SeqMatch 17; Missmatches 87;  
 ; Indexes 51;  
 ; Gaps 10;

3 GGLKRKHSDELLEEEERNEWSPAGLQSYQQALLRISLDKVQRSLGPRAPSILRRHVLIHNTL 62  
 659 GGLK--SPWRGEYKEPRHPPENQANYHQ----TLHAQPRELSPRAPGPRPAEETTGRA 710

63 QQLQAAALR LAPAPALPPE----PLFLGEEEDFSLSATIGSILRELDTSMDGTTEPPQN PVT 117  
 711 PKLQ----PPLPPEPKNSPPPLTLSKEE----SGMWEPLPLSSLEPAPARN PSS 758

118 PLGLQNEVPPQ----PDPVFEALSSRY----LGDSGLDDFFLDIDTSAVEKEPA 164  
 759 P-ERKATVPEQELQQLEIEFLNSLSQFSLEEQILSCLSIDSLSDDS---EKNPS 814

165 RAPPEPHNLFCAPGSW----EWN 184  
 815 KASQSSRDTLSSGVHSWSQAEEARSSSWN 843

RESULT 9  
 Sequence 1, Application US/09257703-1  
 Patent No. 6265538  
 GENERAL INFORMATION:  
 APPLICANT: Greene, Warner C.  
 APPLICANT: Lin, Xin  
 APPLICANT: Gelezunias, Romas  
 TITLE OF INVENTION: A NOVEL INHIBITOR OF THE INFAMMATORY RESPONSE INDUCED  
 BY TNF-ALPHA AND IL-1  
 FILE REFERENCE: 30448.61USU1  
 CURRENT APPLICATION NUMBER: US/09/257,703  
 CURRENT FILING DATE: 1999-02-25  
 EARLIER APPLICATION NUMBER: 60/076,299  
 EARLIER FILING DATE: 1998-02-27  
 NUMBER OF SEQ ID NOS: 2  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 1  
 LENGTH: 947  
 TYPE: PRT  
 ORGANISM: NF-KAPPA B INDUCING KINASE (NIK)  
 US-09-257-703-1

Query Match 8.8%; Score 90.5; DB 3; Length 947;  
 Best Local Similarity 25.8%; Pred. No. 4;  
 Matches 54; Conservative 17; Mismatches 87; Indels 51; Gaps 10;

3 GGLKRKHSDELLEEEERNEWSPAGLQSYQQALLRISLDKVQRSLGPRAPSILRRHVLIHNTL 62  
 659 GGLK--SPWRGEYKEPRHPPENQANYHQ----TLHAQPRELSPRAPGPRPAEETTGRA 710

63 QQLQAAALR LAPAPALPPE----PLFLGEEEDFSLSATIGSILRELDTSMDGTTEPPQN PVT 117  
 711 PKLQ----PPLPPEPKNSPPPLTLSKEE----SGMWEPLPLSSLEPAPARN PSS 758

118 PLGLQNEVPPQ----PDPVFEALSSRY----LGDSGLDDFFLDIDTSAVEKEPA 164  
 759 P-ERKATVPEQELQQLEIEFLNSLSQFSLEEQILSCLSIDSLSDDS---EKNPS 814

165 RAPPEPHNLFCAPGSW----EWN 184  
 815 KASQSSRDTLSSGVHSWSQAEEARSSSWN 843

RESULT 11  
 US-08-778-717-9  
 Sequence 9, Application US/087778717  
 Patent No. 6602689  
 GENERAL INFORMATION:  
 APPLICANT: Ueno, Eiichi  
 APPLICANT: Nobuyuki, Fujii  
 APPLICANT: Okada, Masahisa  
 TITLE OF INVENTION: FUSED PROTEIN  
 TITLE OF INVENTION: FUSED DNA SEQUENCE, FUSED PROTEIN  
 TITLE OF INVENTION: EXPRESSED FROM SAID FUSED DNA SEQUENCE AND METHOD FOR  
 NUMBER OF SEQUENCES: 21  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
 STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400  
 CITY: ARLINGTON  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22202  
 COMPUTER READABLE FORM:  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/778,717  
 FILING DATE: 12-DEC-1996  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: JP 3522225/1995  
 FILING DATE: 28-DEC-1995  
 ATTORNEY/AGENT INFORMATION:  
 NAME: OBLON, NORMAN F.  
 REGISTRATION NUMBER: 24,618  
 REFERENCE/DOCKET NUMBER: 2084-031-0  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220  
 INFORMATION FOR SEQ ID NO: 9:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 215 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 ORIGINAL SOURCE:

ORGANISM: RECOMBINANT  
 PUBLICATION INFORMATION:  
 AUTHORS: NOBUTUKI FUJII ET AL,  
 TITLE: FUSED DNA SEQUENCE, FUSED PROTEIN EXPRESSED  
 FROM SAID FUSED DNA SEQUENCE AND METHOD OF  
 EXPRESSING SAID FUSED PROTEIN  
 TITLE: RELEVANT RESIDUES IN SEQ ID NO: 9; FROM 1 TO 215  
 US-08-778-717-9

Query Match 8.7%; Score 89.5; DB 4; Length 215;  
 Best Local Similarity 23.7%; Pred. No. 0.72; Matches 41; Indels 41; Gaps 6;  
 Matches 41; Conservative 24; Mismatches 67; Indels 41; Gaps 6;

Qy 2 EGGLKRKHSKDLLEE--ERNEWSPAGLQSYYQQLLRLRISLDKVQRLSLG-  
 Db 31 EGKAQPKVIEVIEDELYNCAKEAMEACPVSAITIEAGGSSLVPRGSEEFMCQIHGLSPTP 90

Qy 47 -PRAPSLLRHVLINTLQQLQAAIRLAPAPA-----LPPEPLFLGEEDFSLS 92  
 Db 91 IPKAP--RGLSTTHWLNFQAYRLQPRPSDFDFQQLRRFLKLALKTPIWLNPIDYSL 147

Qy 93 ATI----GSILRELDTSMDGTEPPQNPNVTPLG---LQNEVPQPQDPVPLEA 136  
 Db 148 ASLIPKGYPGRVVEILNLVKRNQVSPSAPAAPVPTPICPTTTPPPPPPSPEA 200

RESULT 12  
 US-09-849-602-18  
 Sequence 18, Application US/09849602  
 Patent No. 6794501

GENERAL INFORMATION:  
 APPLICANT: Scanlan, Matthew J.  
 APPLICANT: Old, Lloyd J.  
 APPLICANT: Stockert, Elisabeth  
 APPLICANT: Chen, Yao-Tseng  
 TITLE OF INVENTION: Colon Cancer Antigen Panel  
 FILE REFERENCE: L0461/7105 (JVR)  
 CURRENT APPLICATION NUMBER: US/09/849,602  
 CURRENT FILING DATE: 2001-05-04  
 NUMBER OF SEQ ID NOS: 30  
 SOFTWARE: PatentIn version 3.0  
 SEQ ID NO 18  
 LENGTH: 897

TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-849-602-18

Query Match 8.6%; Score 88.5; DB 4; Length 897;  
 Best Local Similarity 24.4%; Pred. No. 5.7;  
 Matches 50; Conservative 23; Mismatches 75; Indels 57; Gaps 9;

Qy 13 EEEERNEWSPAGLQSYYQQLQQLQAAIRLAPAPA-----ALLRISLDKVQRLSLGPRAP---SL 52  
 Db 208 QQEAEER----QALQSLRQGGTLTGKEMSTSSISPGCLLGVALE----GDGSPHGHASL 256

Qy 53 RRHVLINTLQQLQQLQAAIRLAPAPA-----ALLRISLDKVQRLSLGPRAP---SL 52  
 Db 257 LQHVLL--LEQARQOSTLIAWPLHGQSPLVGERVATSMRTVGKLPRHRPLRSRTQSSP- 312

Qy 113 QNPVTPPLGLQNEVPQPQDPVFLFELDIDTSAVEKEPARAPPEP 170  
 Db 313 -LPOSQPAOLQQLVMQQHQOFLEKOKQQQLQLGK----ILTKTGEPLRQPTTHPEET 364

Qy 171 PHNLFCAPGSWENELDHIMEIILG 195

RESULT 14  
 US-10-118-328-4  
 Sequence 4, Application US/10118328  
 Patent No. 6773904

GENERAL INFORMATION:  
 APPLICANT: YAN, Chunhua et al.  
 TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
 NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
 PROTEINS, AND USES THEREOF  
 FILE REFERENCE: C10C1220

CURRENT APPLICATION NUMBER: US/10/118,328  
 CURRENT FILING DATE: 2002-04-09  
 PRIOR APPLICATION NUMBER: 60/282,460  
 PRIOR FILING DATE: 2001-04-10  
 NUMBER OF SEQ ID NOS: 5  
 SOFTWARE: FastSEQ for Windows Version 4.0  
 SEQ ID NO 4  
 LENGTH: 709

TYPE: PRT  
 ORGANISM: Mus musculus  
 US-10-118-328-4

Query Match 8.5%; Score 87.5; DB 4; Length 709;  
 Best Local Similarity 26.6%; Pred. No. 5.3;

Matches 46; Conservative 18; Mismatches 62; Indels 47; Gaps 9;  
 Qy 5 LKRRHSDLEEEERNEWSPAAGLQSYQQALLRISLD-----KVQRSLG----- 46  
 Db 440 LKGNLINFEKKRKEWE-ILARIQQLQRCQRYSLSPRPPIALAQRQOLSEEQSYRVSR 498  
 Qy 47 -----PRAPSLRRHVLIHTNLLQQLQQAALRLAPAPA----- 47  
 Db 499 VIEPPAASCPSPRIRRISLTKRLSAKLREKNSPGGSPGDPSSP-----TSSVS 550  
 Qy 95 IG---SILRELDTSMDGTETP----PQNPTVPLGLQNEVPPQDPVFEALSSR 140  
 Db 551 PGSPSSSPRNREPPPGSPPASPGQSPSTKLSLMD-PPGPWVvRLTPSSSR 602

## RESULT 15

US-08-259-451-5

; Sequence 5, Application US/082599451

; Patent No. 6406841

; GENERAL INFORMATION:

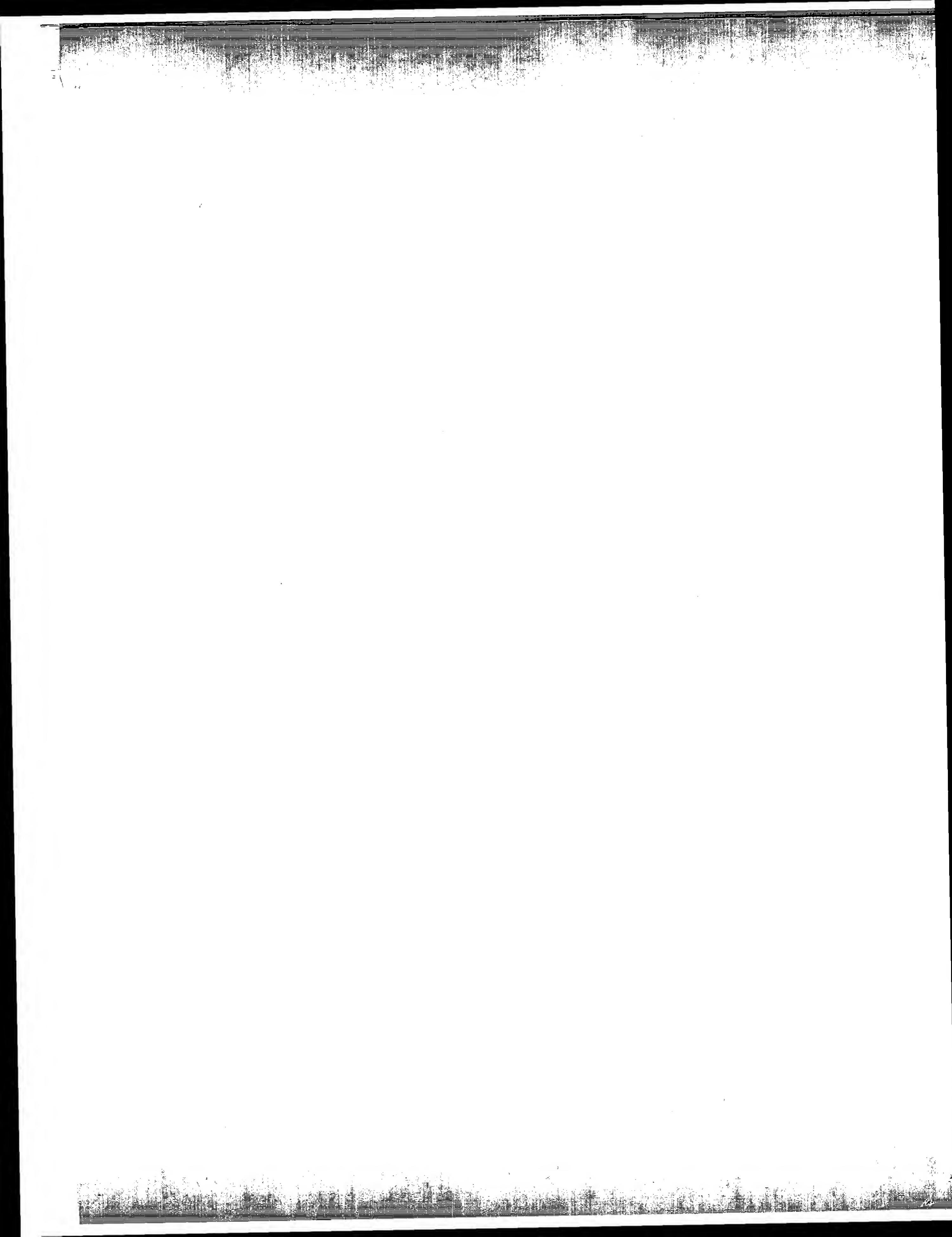
; APPLICANT: Lee, Helen H.  
 ; APPLICANT: Swanson, Priscilla A.  
 ; APPLICANT: Idler, Kenneth B.  
 ; APPLICANT: Rosenblatt, Joseph D.  
 ; APPLICANT: Chen, Irvin S. Y.  
 ; APPLICANT: Golde, David W.  
 ; APPLICANT: Robertson, Eugene F.  
 ; APPLICANT: Stephens, John E.  
 ; APPLICANT: Chan, Emerson W.  
 ; APPLICANT: Buytendorp, Mark H.  
 ; APPLICANT: Johnson, Joan E.  
 ; APPLICANT: Motley, Cheryl T.  
 ; APPLICANT: Peterson, Bryan  
 ; APPLICANT: Edwards, Michelle  
 ; APPLICANT: Guidinger, Peggy  
 ; APPLICANT: Tate, Cynthia

; TITLE OF INVENTION: HTLV-IINRA Compositions  
 ; TITLE OF INVENTION: and Assays for Detecting HTLV Infection  
 ; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Abbott Laboratories  
 ; STREET: One Abbott Park Road  
 ; CITY: Abbott Park  
 ; STATE: Illinois  
 ; COUNTRY: USA  
 ; ZIP: 60064

; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
 ; MEDIUM TYPE: storage  
 ; COMPUTER: IBM  
 ; OPERATING SYSTEM: MS-DOS  
 ; SOFTWARE: WordPerfect 5.1  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/259,451  
 ; FILING DATE: 20-JUN-1994  
 ; CLASSIFICATION: 436  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/086,415  
 ; FILING DATE: 01-JUL-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Daniel W. Collins  
 ; REGISTRATION NUMBER: 31,912  
 ; REFERENCE/DOCKET NUMBER: 5381.US.P1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (708) 937-6365  
 ; INFORMATION FOR SEQ ID NO: 5:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 136 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS:  
 ; TOPOLOGY: unknown  
 ; US-08-259-451-5





Rajandream, R.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.  
Submitted to the EMBL Data Library, September 1998  
A; Reference number: Z21933  
A; Accession: T40514  
A; Status: preliminary; translated from GB/EMBL/DDBJ  
B; Molecule type: DNA  
A; Cross-references: UNIPROT:O74402; EMBL:AL031534; PIDN:CAA20737.1; GSPDB:GN00067; SPDB:  
A; Experimental source: strain 972h-; cosmid C4F6  
C; Genetics:

A; Gene: SPDB:SPBC4F6.17C  
A; Map position: 2  
C; Superfamily: endopeptidase Clp ATP-binding chain  
C; Keywords: ATP; molecular chaperone; nucleotide binding; P-loop  
F: 139-146/Region: nucleotide-binding motif A (P-loop)  
E: 207-212/Region: nucleotide-binding motif B  
E: 539-546/Region: nucleotide-binding motif A (P-loop)  
E: 607-612/Region: nucleotide-binding motif B  
F: 145/Binding site: ATP (Lys) #status predicted  
F: 545/Binding site: ATP (Lys) #status predicted

Query Match	9.4%;	Score 97;	DB 2;	Length 803;
Best Local Similarity	24.2%;	Pred. No. 2.8;		
Matches	53;	Conservative	31;	Mismatches 73; Indels 62; Gaps 11;

```

5 LKRKHSDLEEEER----WEWSAGLQSYQQAA----LLRISLDKVQRSLGPRA----- 49
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
377 LESKLTDLKEEQDRKLSAAWEERKLDSIKRAKTELEQARIELERTQRE-GNYARASELQ 435
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
50 ----PSLRRHV----LIHTNLQQLQAAQLR LAPAPALPPEPLFLGEED--FS 90
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
436 YAIIPPELERSVPKEEKTLBEEKKPMSMVDSTSVDIAVVVS RATGIPPTNLMGERDKLLN 495
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
91 LSATIGSILRELDTSMDGTETPPQNTPVTPLGLQNEVPPQDPVFILEALSSRYLGDSG---- 146
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
496 MEQTIGKKIIGQDEALKAIADAVR-LSRAGLQNTNRPP-----LASFLFLGPTGVGKT 546
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
147 ----LDDFFELDIDTSAV----EKEP-ARAPPEP 171
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
547 ALTKALAEEFLFDTDKAMIRFDMMSEFQERHTIARLIGSPP 585
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 3  
G8641 Unknown protein [imported] - Arabidopsis thaliana  
C; Species: Arabidopsis thaliana (mouse-ear cress)  
C; Date: 02-Mar-2001 #sequence\_change 09-Jul-2004  
C; Accession: G86442  
B; Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, J.S.; Creasy, T.H.; Dewar, K.; Hansen, N.F.; Hughes, B.; Huizar, L.  
A; Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lurov, J.S.; Maiti, R.; Marziali, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A; Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A; Title: Sequence and analysis of chromosome 1 of the Plant Arabidopsis.  
A; Reference number: A86141; MUID:21016719; PMID:11130712  
A; Accession: G86441  
A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-1201 <STO>  
A; Cross-references: UNIPROT:Q9C6S1; GB:AE005172; PIDN:911136725; PIDN:AAG31306.1; GSPDB:G  
C; Genetics:

A; Map position: 1  
C; Map position: 2  
C; Map position: 3  
C; Map position: 4  
C; Map position: 5  
C; Map position: 6  
C; Map position: 7  
C; Map position: 8  
C; Map position: 9  
C; Map position: 10  
C; Map position: 11  
C; Map position: 12  
C; Map position: 13  
C; Map position: 14  
C; Map position: 15  
C; Map position: 16  
C; Map position: 17  
C; Map position: 18  
C; Map position: 19  
C; Map position: 20  
C; Map position: 21  
C; Map position: 22  
C; Map position: 23  
C; Map position: 24  
C; Map position: 25  
C; Map position: 26  
C; Map position: 27  
C; Map position: 28  
C; Map position: 29  
C; Map position: 30  
C; Map position: 31  
C; Map position: 32  
C; Map position: 33  
C; Map position: 34  
C; Map position: 35  
C; Map position: 36  
C; Map position: 37  
C; Map position: 38  
C; Map position: 39  
C; Map position: 40  
C; Map position: 41  
C; Map position: 42  
C; Map position: 43  
C; Map position: 44  
C; Map position: 45  
C; Map position: 46  
C; Map position: 47  
C; Map position: 48  
C; Map position: 49  
C; Map position: 50  
C; Map position: 51  
C; Map position: 52  
C; Map position: 53  
C; Map position: 54  
C; Map position: 55  
C; Map position: 56  
C; Map position: 57  
C; Map position: 58  
C; Map position: 59  
C; Map position: 60  
C; Map position: 61  
C; Map position: 62  
C; Map position: 63  
C; Map position: 64  
C; Map position: 65  
C; Map position: 66  
C; Map position: 67  
C; Map position: 68  
C; Map position: 69  
C; Map position: 70  
C; Map position: 71  
C; Map position: 72  
C; Map position: 73  
C; Map position: 74  
C; Map position: 75  
C; Map position: 76  
C; Map position: 77  
C; Map position: 78  
C; Map position: 79  
C; Map position: 80  
C; Map position: 81  
C; Map position: 82  
C; Map position: 83  
C; Map position: 84  
C; Map position: 85  
C; Map position: 86  
C; Map position: 87  
C; Map position: 88  
C; Map position: 89  
C; Map position: 90  
C; Map position: 91  
C; Map position: 92  
C; Map position: 93  
C; Map position: 94  
C; Map position: 95  
C; Map position: 96  
C; Map position: 97  
C; Map position: 98  
C; Map position: 99  
C; Map position: 100  
C; Map position: 101  
C; Map position: 102  
C; Map position: 103  
C; Map position: 104  
C; Map position: 105  
C; Map position: 106  
C; Map position: 107  
C; Map position: 108  
C; Map position: 109  
C; Map position: 110  
C; Map position: 111  
C; Map position: 112  
C; Map position: 113  
C; Map position: 114  
C; Map position: 115  
C; Map position: 116  
C; Map position: 117  
C; Map position: 118  
C; Map position: 119  
C; Map position: 120  
C; Map position: 121  
C; Map position: 122  
C; Map position: 123  
C; Map position: 124  
C; Map position: 125  
C; Map position: 126  
C; Map position: 127  
C; Map position: 128  
C; Map position: 129  
C; Map position: 130  
C; Map position: 131  
C; Map position: 132  
C; Map position: 133  
C; Map position: 134  
C; Map position: 135  
C; Map position: 136  
C; Map position: 137  
C; Map position: 138  
C; Map position: 139  
C; Map position: 140  
C; Map position: 141  
C; Map position: 142

RESULT 5

A42029 transcription Factor E3 - mouse

C;Species: Mus musculus (house mouse)

C;Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004

C;Accession: A42029

R;Roman, C.; Matera, A.G.; Cooper, C.; Artandi, S.; Blain, S.; Ward, D.C.; Calame, K.

Mol. Cell. Biol. 12, 817-827, 1992

A;Title: mTFE3, an X-linked transcriptional activator containing basic helix-loop-helix

A;Reference number: A42029; MUID:92123207; PMID:1732746

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-446 <ROM>

A;Cross-references: UNIPROT:Q64092; PIDN:AAB21130.1; PID:g243440

A;Note: sequence extracted from NCBI backbone (NCBIN:76673, NCBIP:76674); A;Reference number: A42042; MUID:85216449; PMID:2582407

A;Molecule type: DNA

A;Residues: 1-433 <SH1>

A;Cross-references: UNIPROT:P03346; GB:MI0060; NID:9329559; PIDN:AB59884.1; PID:g32956;

A;Note: the authors translated the codon TAC for residues 197 and 249 as Thr

FOLJH2

gag polyprotein - human T-cell lymphotropic virus type 2

N;Alternate names: core polyprotein

N;Contains: core protein p12; core protein p15; core protein p24

C;Species: human T-cell lymphotropic virus type 2, HTLV-2

A;Note: host Homo sapiens (man)

C;Date: 28-Feb-1986 #sequence\_revision 28-Feb-1986 #text\_change 09-Jul-2004

C;Accession: A03944

R;Shimotohno, K.; Takahashi, Y.; Shimizu, N.; Gojobori, T.; Golde, D.W.; Chen, I.S.Y.; Proc. Natl. Acad. Sci. U.S.A. 82, 3101-3105, 1985

A;Title: Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus

A;Reference number: A94042; MUID:85216449; PMID:2582407

A;Accession: A03944

Query Match 9.2%; Score 95.5; DB 2; Length 446;

Best Local Similarity 27.3%; Pred. No. 1.7;

Matches 48; Conservative 24; Mismatches 61; Indels 43; Gaps 8;

QY 12 LEEEEEWPSAGLQSYQQALLRISLDKVQRSLGPRAPSLLRHVLIHNTLQQQLQAALRL 71

Db 272 LQKECQR----SKDLESRQR----SLEQANRSLQLRIQEL----ELQAQIHG 311

QY 72 APAPALPPEPLFLGEEEDFSLSATIGSILIRELDTSMDGTTEPPQNPTVPLGLQNEVPPQDPD 131

Db 312 LP---VPPNPGLLSLTSSVSDSLKP--EQLDIEEGRSTTFHVSGGPAQNAPPQPPA 366

Query Match 9.0%; Score 93; DB 1; Length 433;

Best Local Similarity 23.1%; Pred. No. 2.7;

Matches 40; Conservative 24; Mismatches 67; Indels 42; Gaps 7;

QY 47 PRAPSLRRHVLIHNTLQQQLQAALRLAPAPA-----LPPEPLFLGEEEDFSLSA 93

Db 13 PKAP---RGLSTHHLNLFLQAYRLQPRPSDFDFQQLRFLKLAALKTPIWLNPIDSLAA 69

QY 94 TI-----GSTLRELDTSMDGTTEPPQNPTVPLG---LQNEVPPQDPVFLEA-LSSRYLGA 143

Db 70 SLIPKGYPGRVVEIIINILVRQNQVSNSPAAAPVPTPICPCTTTPPPPPSPPEAKVPPY-- 127

QY 144 DSGLDDFFLDIDTSMDGTTEPPQNPTVPLGLQNEVPPQDPD 196

Db 128 -----VEPTTQCFPILHFGAP--SAHRPWNMKDLQAIKQEVS 166

RESULT 6

A54602 microtubule-associated serine/threonine protein kinase MAST205 - mouse

C;Species: Mus musculus (house mouse)

C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004

C;Accession: A54602

R;Walden, P.D.; Cowan, N.J.

Mol. Cell. Biol. 13, 7625-7635, 1993

A;Title: A novel 205-kilodalton testis-specific serine/threonine protein kinase associated with a protein kinase homology domain: protein kinase ATP-binding motif E

A;Reference number: A54602; MUID:94067123; PMID:8246979

A;Accession: A54602

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1734 <WAL>

A;Cross-references: UNIPROT:Q60592; GB:U02313; NID:9406057; PIDN: AAC04312.1; PID:g406058

C;Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homologous

C;Keywords: ATP F;451-726/Domain: protein kinase homology <KIN>

F;459-467/Region: protein kinase ATP-binding motif E

Query Match 9.1%; Score 93.5; DB 2; Length 1734;

Best Local Similarity 26.3%; Pred. No. 15;

Matches 51; Conservative 23; Mismatches 71; Indels 49; Gaps 8;

QY 25 GLQSYQQAL---LRISLDKVQRSLGPRAPSLLRHVLIHNTLQQQLQAALRLAPALPP 79

Db 1308 GSQSFPTKLHLSPLGRLSPKSAEPPERSPLKR-----VQSAEKLAALAAA 1357

QY 80 EPLFLGEEDFSLSATIGSILIRELDTSMDGTTEPPQNPTVPLGLQNEVPPQDPVFLEA 136

Db 1358 KRL-APSRKHSDLPLHGEKLKEL-----TPREASPLEVVGTRSVLSGKGPLPGKGVLQP 1410

RESULT 8

A82255 hypothetical protein VC0998 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C;Species: Vibrio cholerae

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004

C;Accession: A82255

R;Heidelberg, J.P.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; chardson, D.; Birmolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, I., R.R.; Mekalanos, J.J.; Ventner, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A;Title: Sequence of both chromosomes of the cholera pathogen Vibrio cholerae

A;Reference number: A82035; MUID:10952301

A;Accession: A82255

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-1621 <HEI>

A;Cross-references: UNIPROT:Q9KTA5; GB:AE004181; PIDN:AB003852; NID:99655454; PIDN:AAF9415

A;Experimental source: serogroup O1; strain N16961; biotype El Tor

C;Genetics:

A;Gene: VC0998

A; Map position: 1

Query Match 8.9%; Score 91.5; DB 2; Length 1621;  
Best Local Similarity 27.4%; Pred. No. 20;  
Matches 49; Conservative 22; Mismatches 63; Indels 45; Gaps 8;

Qy 13 EEEERWENSPAGLQSYYQALLRISLDKVQRSLGPRAPSLLRRHVLIHNTLQQLQAALRLA 72  
Db 530 EEDDEFDLGAGVAGDQ----DLDLFAFASIEQA-----DLEQLEA--KAI 570  
Qy 73 PAPALPFPFLFGEEDFSLSATIGSILRELDTSMDGTE-----PPQNPNPVTPL- 119  
Db 571 DETALLE--I LAEQDAPLSEESTELLDLLDDFDKEPENDEFDAQTADLLOPEEPILDLE 628  
Qy 120 ---GLONEVPQPDVFLEA----LSSRYLGDSGLDDFFLDIDTSAVBKEPARAPPE 169  
Db 629 EDSTQNLNEVILGEPVPEELASGLEIDQNSTEMELLDLDD--LDLDESIBATEFSVAPE 685  
Qy 135 EA LSSRYLGDSGLDDFFLDIDTSAVEKEPARAPPE 170  
Db 433 IASTGQQLPTGGFSG----SGPPPISQQLVLQPPPSP 463

RESULT 9

AG0565 probable membrane protein STY0554 [Imported] - *Salmonella enterica* subsp. *enterica* serovar Typhi C; Species: *Salmonella enterica* subsp. *enterica* serovar Typhi Note: this species has also been called *Salmonella typhi* C; Date: 09-Nov-2001 #sequence\_revision 09-Nov-2002 C; Accession: AG0565 R; Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, S.; T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moule, S.; O'Gaora, P. Nature 413, 848-852, 2001 A; Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A; Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* serovar Typhi C; Reference number: AB0502; PMID:21534947; PMID:11677608 A; Accession: AG0565 A; Status: preliminary A; Cross-references: GB:AL513382; PIDN:CAD04993.1; PID:916501778; GSPDB:GN00176 A; Residues: 1-804 <PAR> A; Molecule type: DNA C; Genetics:

A; Gene: STY0554 C; Superfamily: Escherichia coli probable membrane protein ybbP

Query Match 8.8%; Score 90.5; DB 2; Length 804;  
Best Local Similarity 28.7%; Pred. No. 9.9;  
Matches 37; Conservative 16; Mismatches 39; Indels 37; Gaps 7;

Qy 38 LDKVQRSLGPRAPS-----LRRHVLIHNTLQQLQAALRLAPALPPE 80  
Db 478 LDRWQQQLPPSPNPFINTIASEQVAPLKAFLAEHQVIPQTFYPIVRA-RLTEINGNPTE 536  
Qy 81 PLFLGEEEDFSLSATIGSILRELDTSMDGTEPPQNPNPVTPLGLQNEVPPQDPVFL-EALS 139  
Db 537 ---GQODESLN-----RELNLTWQDTRPAHNPL---VAGHWWPKPGEVSMEEGLAK 581  
Qy 140 RY---LGDS 145  
Db 582 RLNVKLGDS 590

RESULT 10

T42644 hypothetical protein DKFZP56N1047.1 - human (Fragment)  
C; Species: Homo sapiens (man)  
C; Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jul-2004  
C; Accession: T42644 R; Ottenwaelder, B.-i; Obermaier, B.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S. Submitted to the Protein Sequence Database, November 1999  
A; Reference number: Z22231 A; Accession: T42644 A; Status: preliminary A; Molecule type: mRNA  
A; Residues: 1-651 <AAA> A; Cross-references: UNIPROT:Q9UF93; EMBL:All33109

A; Experimental source: fetal kidney; clone DKFZP56N1047  
C; Genetics:  
A; Note: DKFZP56N1047.1

Query Match 8.5%; Score 88; DB 2; Length 651;  
Best Local Similarity 24.4%; Pred. No. 12;  
Matches 38; Conservative 26; Mismatches 68; Indels 24; Gaps 7;

Qy 27 QSYQQALLRISLDKVQRSLGPRAPSLLRRHVLIHNTLQQLQAALRLA-PAPALPPEPL-- 82  
Db 320 QSCQQPPQQPSPSPQQVPPQPM-AGPLVLTQSVQGIQASSQSQVQYPAVSFPPQHLLP 378  
Qy 83 ----FLGEEEDFSLSATIGSILRELDTSMDGTEPPQNPNPVTPLGLQNEVPPQDPVFL- 134  
Db 379 VSPTQHFPMRDD--VATQFGQMTLSRQSSGETPEPPSGPVYPSSLM----PQPAQQPSYV 432  
Qy 135 EA LSSRYLGDSGLDDFFLDIDTSAVEKEPARAPPE 170  
Db 433 IASTGQQLPTGGFSG----SGPPPISQQLVLQPPPSP 463

RESULT 11

T26998 hypothetical protein Y48B6A.6 - *Caenorhabditis elegans*  
C; Species: *Caenorhabditis elegans*  
C; Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C; Accession: T26998 R; Wall, M. Submitted to the EMBL Data Library, September 1999  
A; Reference number: Z20297 A; Accession: T26998 A; Status: preliminary; translated from GB/EMBL/DDBJ  
A; Molecule type: DNA  
A; Residues: 1-716 <WIL> A; Cross-references: UNIPROT:Q9U2A6; EMBL:AL110490; NID:e1542263; PIDN:CA554442; 1<CESE> A; Experimental source: clone Y48B6A C; Genetics:  
A; Gene: CESP; Y48B6A.6 A; Introns: 38/3; 196/3; 437/2; 460/3; 518/1; 617/3; 673/3  
A; Molecule type: DNA  
A; Cross-references: GB:AL513382; PIDN:CAD04993.1; PID:916501778; GSPDB:GN00176  
C; Genetics:

Query Match 8.5%; Score 88; DB 2; Length 716;  
Best Local Similarity 25.0%; Pred. No. 14;  
Matches 49; Conservative 29; Mismatches 80; Indels 38; Gaps 9;

Qy 7 RKHSDEEEERRNEWS--PAGLQSYQQALLRISLDKVQ---RSLGPRAPSLLRRHVLIHNT 61  
Db 219 RSDSIEEEERRKESETASFEELLEAEIMRISRSVPPVLSIPPVPLPPPP--NIPLPTI 275  
Qy 62 LQLOQAAHLRAPALPPEPLGEEEDFSLSATIGSII-----LRELD 103  
Db 276 PQEVQSPSPSPRPTSVPPPIPSPGPPSEDVNMDLIESFSDSVIFNNMSMSPPPPLPPLRE-- 333  
Qy 104 TSMGDGTE-PPQNPVTPLGLQNEVPPQDPVFL-SRYLGDSGLDDFFLDIDTSAVEK 161  
Db 334 SSLETLEVTPEDPVTESKV--EASSTPLPKATESLNESSIKALEGLEYVKALEAQEASDDR 391  
Qy 162 EPARAP----PEPP 171  
Db 392 PSAPTPIRDSLPPPP 407

RESULT 12

A56508 anucleate primary sterigmata A (apsA) protein - *Emericella nidulans*  
C; Species: *Emericella nidulans*, *Aspergillus nidulans*  
C; Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 09-Jul-2004  
C; Accession: A56508 R; Fischer, R.; Timberlake, W.E. Submitted to the Protein Sequence Database, November 1999  
A; Title: Aspergillus nidulans apsA (anucleate primary sterigmata) encodes a coiled-coil protein  
A; Reference number: A56508; PMID:95164553; PMID:7860626  
A; Accession: A56508  
A; Status: preliminary

A; Molecule type: DNA  
A; Residues: 1-1676 <FIS>  
A; Cross-references: UNIPROT:Q00083; GB:X82289; NID:9683499; PID:9683500  
C; Genetics:  
A; Gene: apsa  
A; Introns: 149/3

Query Match 8.5%; Score 88; DB 2; Length 1676;  
Best Local Similarity 20.5%; Pred. No. 42;  
Matches 32; Conservative 19; Mismatches 51; Indels 54; Gaps 6;

Qy 36 ISLDKVQRSLGPRAPSILRHHVLIHNTLQQALRAPALPPEPLFLGEEDFSLSSATI 95  
Db 837 VSFEETE-PVAPSFFPELRTAFFVGSTTEPVAAAPVVPPEVALSP-----I 880

Qy 96 GSILRELDTSMDGTEPPQNPTPLGLQNEVPPQDPVFL-----  
Db 881 SS-----QTQQTPEVIP-----APPEPEIYVPEMAFSQILVEDTLPILAKLP 924

Qy 135 EALSSRYLGDSGLDDFFFLDIDTSAVEKEPARAPPEP 170  
Db 925 EPAPERVFAEQGTSTDIADVSVAISSEQTE-PVEP 959

RESULT 13

A34596 transcription factor E3 - human (fragment)  
C; Species: Homo sapiens (man)  
C; Date: 06-Jul-1990 #sequence\_revision 09-Oct-1992 #text\_change 09-Jul-2004  
C; Accession: A34596; S10379  
R; Beckmann, H.; Su, L.K.; Kadesch, T.  
Genes Dev. 4, 167-179, 1990  
A; Title: TFE3: a helix-loop-helix protein that activates transcription through the immun  
A; Reference number: A34596; MUID:90249724; PMID:2338243  
A; Accession: A34596  
A; Status: preliminary  
A; Molecule type: mRNA  
A; Residues: 1-536 <BER>  
A; Cross-references: UNIPROT:P19532; EMBL:X51330; NID:g37061; PID:g13353  
A; Note: the authors translated the codon ACC for residue 433 as Ser, and GAG for residue  
C; Genetics:  
A; Gene: GDB:TFE3  
A; Cross-references: GDB:125870; OMIM:314310  
A; Map position: Xp11.23-Xp11.22  
C; Keywords: DNA binding; transcription factor

Query Match 8.4%; Score 86.5; DB 2; Length 536;  
Best Local Similarity 25.4%; Pred. No. 13; Mismatches 55; Indels 59; Gaps 10;

Qy 12 LEEEFERNEWSPAGLQSYQQALLRISLDKVQRSLGPRAPSILRHHVLIHNTLQQQLQAAARI 71  
Db 192 LQKEQQR----SKDLESRQR----SLEQANRSLQRQEL----ELQAQIHG 231

Qy 72 APAPALPPEPLFLGEEDFSLSATISILRE-LDTSMDGTE--PPQNPVTPLG 120  
Db 232 LPVPGTP----GLLSLATSTSDSLKPEQLDIEEGRPGARTFHVGGAQN---- 279

Qy 121 LQNEVPQPQ----DPVFLREALSRYLGDSGLDDFFFLDIDTSAVEKE--PAR 165  
Db 280 APHQQQPAPPSSDALLDLHFPSDHLDLG-DPEHLGLEDILMEEEEEVGVGGALSPLR 338

Qy 166 APPEP 170  
Db 339 AASDP 343

RESULT 14

T14355 protein-tyrosine-phosphatase (EC 3.1.3.48) TD14 - rat  
C; Species: Rattus norvegicus (Norway rat)  
C; Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C; Accession: T14355

Query Match 8.4%; Score 86.5; DB 2; Length 1494;  
Best Local Similarity 29.1%; Pred. No. 48;  
Matches 37; Conservative 19; Mismatches 40; Indels 31; Gaps 9;

Qy 62 IQQOQAALRLAPAPALPPEPLFLGEEDFSLSATISILRE----LDTSMDGTEPPQNP 115  
Db 625 LMQPRAAVPMAFPGPVLYPAPVYSE----LAMDVR-PATTIVDWSQAPISSHMALRPGPAPPPQ-P 675

Qy 116 VTPLGLQNENVPBPQ---PDPVFLREALSSRYLGDSGLDDEFLDIDT-SAVEKEPARAPPEP 171  
Db 676 I--VGLPSAPPQFQSGPE---LAMDVR-PATTIVDWSQAPISSHMALRPGPAPPPQ-P 727

Qy 172 HNLFCAP 178  
Db 728 ---CFP 730

RESULT 15

B37761 ntrA protein - Thiobacillus ferrooxidans  
C; Species: Thiobacillus ferrooxidans  
C; Date: 31-May-1991 #sequence\_revision 31-May-1991 #text\_change 09-Jul-2004  
C; Accession: B37761  
R; Berger, D.K.; Woods, D.R.; Rawlings, D.E.  
J. Bacteriol. 172, 4399-4406, 1990  
A; Title: Complementation of Escherichia coli sigma(54) (intra) -dependent formate hydrogen  
A; Reference number: A37761; MUID:90330545; PMID:2198257  
A; Accession: B37761  
A; Status: preliminary  
A; Molecule type: DNA  
A; Residues: 1-475 <BER>  
A; Cross-references: UNIPROT:P24695; GB:M58480; PID:g154642; PID:AAA27379.1;  
C; Superfamily: Pseudomonas transcription initiation factor sigma  
C; Keywords: DNA binding; transcription regulation

Query Match 8.3%; Score 86; DB 2; Length 475;  
Best Local Similarity 21.7%; Pred. No. 12;  
Matches 40; Conservative 31; Mismatches 75; Indels 38; Gaps 6;

Qy 1 MEGGLKRHKHSDELLEEEERNEWSPAGLQSYQQALLRISLDKVQR----SLGPRA--PSLR 53  
Db 307 MAGGKDAAHKYIQDQLNEARWFIKSLQSRQDTILKVARAIVERQKDFFANGPESMRPMVL 366

Qy 54 RHVLIHNTLQQQALRAPALPPEPLFLGEEDFSLSATIG----SILRELDT 104  
Db 367 RH--IADAVENHESTSVSRVTNQKYMITPRGLYEFKYFFSSHVGTDSGGASATAIRALLI 424

Qy 105 SMDGTEPPQNPVTPLGLQNEVPPQDPVFLREALSSRYLGDSGLDDFFLDTSAVEKEPA 164  
Db 425 RMTOQAEDAQHPLS----DAEIARVLAQDQGIQ---IARRTVAKYREAA 464

Qy 165 RAPP 168  
Db 465 NVPP 468

search completed: November 15, 2004, 14:07:21

Mon Nov 15 17:33:07 2004

Job time : 40 secs

us-10-069-386-2.rpr